1968

A procedure for estimating the amount of trait, method and error variance attributable to a measure

Robert Francis Boruch
Iowa State University

 69-4216

BORUCH, Robert Francis, 1942-
A PROCEDURE FOR ESTIMATING THE AMOUNT OF TRAIT, METHOD AND ERROR VARIANCE ATTRIBUTABLE TO A MEASURE.

Iowa State University, Ph.D., 1968
Psychology, general

University Microfilms, Inc., Ann Arbor, Michigan

A PROCEDURE FOR ESTIMATING THE AMOUNT OF TRAIT, METHOD AND ERROR VARIANCE ATTRIBUTABLE TO A MEASURE

by

Robert Francis Boruch

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Psychology

Approved:

Signature was redacted for privacy.

In Charge ~~o~~f Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1968

TABLE OF CONTENTS

I. INTRODUCTION

This dissertation is based essentially on proposals developed by Wolins (1964) for analysis of multitrait-multimethod data. His exposition may be summarized in the following manner. Given the scores of person i, on a measure of trait j, using method k, one may be able to discover an appropriate linear model of the form,

$$Y_{i(jk)} = \beta_j X_i + \beta_{(jk)j\cdot k} X_{ij} + \beta_{(jk)k\cdot j} X_{ik} + \epsilon_{i(jk)} \; .$$

The X's are scores of people on these conceptual variables. The X's are weighted by the β's. The Y's are the observed scores. The β's are made to be standard score regression weights by imposing the restriction that the X's and Y are distributed with mean zero and a variance of unity. One can also impose restrictions of various kinds on the correlations within methods and traits and/or between methods and traits. The problem is estimation of these standard score regression weights, from which one can estimate for each variable the variance attributable to method, trait and error. The procedure can be related to analysis of variance of classificatory models as well as to regression. This aspect is described in following sections. Part of the procedure is based on previous work done by Stanley (1961) and Campbell (1965), described in the review of literature below.

The importance of the objectives suggested above, relative to psychological measurement, is apparent if one considers the areas of application. Industrial or clinical psychologists, concerned with different ways of measuring different attributes of subjects, can direct inquiry toward assessment of biases of the various methods. Specifically, one might want to determine techniques for adjustment of rater's judgements which are biased with respect to ratees and/or traits. The general multitrait-multimethod paradigm has not been applied to experimental psychology data since the emphasis there has not been individual differences. However, if such procedures are applied, analogous methods of considering effects and interactions can be developed. This possibility is considered below in the section on future research.

From the point of view of the general objectives of science, appropriate reduction of this type of data is desirable. Instead of a systematic perusal of a complicated array of numbers, one can appeal to an analytic procedure. The latter procedure, considered in this dissertation, is certain to be more easily interpreted. The solution, in addition, is better defined statistically in terms of accuracy and precision. The method of analysis to be described is more appropriate than prior suggestions insofar as less restrictive statistical assumptions are required. This aspect is also described below.

The general notation which will be adhered to consists of the following.

A. General Notation

Upper case English and Greek letters represent matrices throughout the paper. English letters designate sample observations or functions of observations while Greek letters designate parameters unless otherwise indicated. Lower case letters represent scalars and, when either lower or upper case are subscripted, they represent the scalar elements of a matrix.

The conventions are generally the same as those used by Joreskog (1967b) and Kempthorne (1952).

B. General Definition of the Problem

In order to delineate the problem attacked in this paper and the rationale upon which the proposed solution is based, some definitions and comments are necessary. The latter are taken, in part, from Thurstone (1947) and from Campbell and Fiske (1959).

A _subject_ is generally the person whose responses are observed.

A _trait_ is defined as any potentially observable attribute of the subject.

A _method_ refers to the means used to observed or assess the trait.

A _measure_ refers to the trait-method combination used to obtain an observation on the subject. The observations yielded by a measure are treated as if they vary on an interval scale.

The definition of measure is simply a formalization of the operationist doctrine as given by, say, Kaplan (1964) that what is measured, and how we measure it are determined jointly. Consider, for example, the trait called neuroticism. The trait might be measured by methods such as psychologists' judgements, graded performance on tasks in stressful situations, or questionnaires. Each observation, provided by the measure, is a function of the subject, the trait, and the method used to assess the trait. The following questions may be asked, based on the comments above. What sort of function is this? Given an observation, what can one say about the extent to which the subject, trait or method contribute to the observation? These questions loosely define the problem area attacked in this research. The definition of the measure forms the literal prototype for the statistical models to be considered below. It might be helpful at this point to be a bit more specific about the usefulness of such questions.

Are some methods of measurement better than others with respect to some criteria? For example, is a psychologist's judgement about the degree of a subject's neuroticism "better than" measurement of neuroticism by means of a questionnaire. "Better than" can be interpreted to mean the extent to which the observation contains less extraneous variance as well as being a better indication of the subject's true trait score. The psychologist may be personally biased against a subject, may give careless judgements or may fail to assimilate information which would yield a valid judgement. Consider also whether or not two traits, measured using the same method, can be distinguished reliably from one another. For example, does a measure of neuroticism yield the same results as a measure of emotionality? Establishing traits as distinct entities is a prevalent problem in personality research. Given a group of people on which to make judgements, a psychologist could conceivably rate the individuals exactly the same for both traits. This event would suggest that there is no reliable way to distinguish between the traits. A trait is said to be valid in a convergent sense if several independent methods of measuring that trait are highly correlated. In the example of neuroticism above, if the results of psychologists' judgements, questionnaires and scores on performance of a task are all linearly related, one can infer that the trait has <u>convergent validity</u>. A trait is said to have <u>discriminant validity</u> if correlations among different measures of the same trait are generally higher than correlations among different traits using the different methods. <u>Method bias</u> may be inferred when the correlations between different traits are higher when the same method is used than when different methods are used.

The problem may be examined relative to the use of measurement in all scientific enterprises. Measurement makes possible discriminations between observations relative to some criterion. The criterion in this instance is a theoretical variable, construct, or trait, hypothesized to underly observations. Measurement may provide estimates of the true value of each experimental unit on the hypothetical trait. Further enhancement of the tenability of the true value is possible with several modes or methods of assessment. The concept or trait is considered valid to the extent that all methods yeild the same true score for an experimental unit, the person. In addition, measurement must provide evidence that true scores on one trait are not identical to those scores conceptually different trait. If such evidence is not provided by data, then the conceptualization of two separate traits is not needed to explain the results of an analyses and a simpler conceptualization is sufficient.

II. REVIEW OF LITERATURE RELEVANT TO THE DESIGN

How can one formulate an approach, based on the definitions and comments above, to the questions outlined? The first consideration, of course, is the design.

The basic data must consist of results of measures as defined above. However, more frequently than not, the data analyzed are the correlations among the observations. This point may need additional amplification.

Ordinarily, one may choose to analyze either raw data or some function of the observations. This, of course, is an interesting general problem for statisticians, discussed by Kempthorne (1967), for example. In the context of this dissertation the problem is related to the following facts.

1. Given a set of measures, the psychologist is frequently interested in the assessment of relations among the measures, rather than the observations, per se. That is, observed variable results are arbitrary to the extent that scales can be manipulated without changing inferences based on the results. With respect to the usual scientific criterion of predictability then, one may choose to examine reproducibility of correlations rather than of observed scores. The usefulness of such an approach will be assumed rather than defended in this paper, although alternative methods will be discussed.

2. An approach to the questions outlined has been developed by Campbell and Fiske (1959) and is described in the literature review. The method of analysis has yielded reasonable results, and is based on examination of correlations rather than raw data. The data structure which they suggest is the sort of data usually collected and which is considered in

this paper.

The raw data may be described by the following array, where $\{T_j(j=1,2,3)\}$ represents the jth trait, and $\{M_k(k=1,2,3)\}$ represents the kth method. Y_{ijk} represents the ith observation on trait j and method k.

M_1			M_2			M_3		
T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
Y_{111}	Y_{121}	Y_{131}	Y_{112}	Y_{122}	Y_{132}	Y_{113}	Y_{123}	Y_{133}
Y_{211}	Y_{221}	Y_{231}	Y_{212}	Y_{222}	Y_{232}	Y_{213}	Y_{223}	Y_{233}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Y_{N11}	Y_{N21}	Y_{N31}	Y_{N12}	Y_{N22}	Y_{N32}	Y_{N13}	Y_{N23}	Y_{N33}

The intercorrelations among the observed measures comprise the available observed data and can be displayed in the array called a <u>multitrait-multimethod matrix</u>. The matrix is illustrated in Table 1. The $r_{(jk)(j'k')}$ represents the correlation between the measure jk and the measure $j'k'$ where j designates trait, and k designates method.

Generally, the following nomenclature is useful in describing the matrix.

1. <u>Heterotrait-heteromethod</u> blocks are composed of the elements $r_{(jk)(j'k')}$ where $j \neq j'$ and $k \neq k'$. Inferences concerning individual differences, independent of method and trait, are based on these coefficients. Such individual differences might stem from trait intercorrelations, for example.

2. Validity diagonals are composed of $r_{(jk)(j'k')}$ elements where $j = j'$ and $k \neq k'$. Inferences concerning convergent validity are based on the size of these coefficients. Inferences concerning discriminant validity are based on the size of these coefficients relative to the size of the coefficients previously described.

3. Heterotrait-monomethod triangles are composed of the elements $r_{(jk)(j'k')}$ where $j \neq j'$ and $k = k'$. Method bias is inferred when these coefficients are large relative to the size of the Heterotrait-heteromethod elements.

The classical reference describing the analysis of multitrait-multimethod matrices in determination of convergent and discriminant validity is Campbell and Fiske (1959). The article presents a method of systematic examination of such matrices. The examination is deductive and inferential but not analytic in a statistical sense. The model which determines the examination is implicit rather than explicit. Such an examination is rather complicated, results often ambiguous.

In an effort to clarify and extend this original work, Campbell has attempted to quantify the examination. In an unpublished manuscript, Campbell (1962) investigated an iterative procedure for estimation of factor analytic weights in a linear model for the data. The iterative procedure proceeds from initial estimates of loadings, based on assumptions about the equality of various loadings. Further assumptions are made about the positions of such factor loadings in the factor structure. The location of elements in the structure define the model in the orthogonal case. Campbell's effort was directed toward circumventing limitations which he suggested were common to all factor analytic techniques. Lack of a

Table 1. Multimethod-multitrait matrix for three methods and three traits

		M_1			M_2			M_3		
		T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
	T_1	1.00					(symmetric)			
M_1	T_2	$r_{(21)(11)}$	1.00							
	T_3	$r_{(31)(11)}$	$r_{(31)(21)}$	1.00						
	T_1	$r_{(12)(11)}$	$r_{(12)(21)}$	$r_{(12)(31)}$	1.00					
M_2	T_2	$r_{(22)(11)}$	$r_{(22)(21)}$	$r_{(22)(31)}$	$r_{(22)(12)}$	1.00				
	T_3	$r_{(32)(11)}$	$r_{(32)(21)}$	$r_{(32)(31)}$	$r_{(32)(12)}$	$r_{(32)(22)}$	1.00			
	T_1	$r_{(13)(11)}$	$r_{(13)(21)}$	$r_{(13)(31)}$	$r_{(13)(12)}$	$r_{(23)(22)}$	$r_{(13)(32)}$	1.00		
M_3	T_2	$r_{(23)(11)}$	$r_{(23)(21)}$	$r_{(23)(31)}$	$r_{(23)(12)}$	$r_{(23)(22)}$	$r_{(23)(32)}$	$r_{(23)(13)}$	1.00	
	T_3	$r_{(33)(11)}$	$r_{(33)(21)}$	$r_{(33)(31)}$	$r_{(33)(12)}$	$r_{(33)(22)}$	$r_{(33)(32)}$	$r_{(33)(13)}$	$r_{(23)(23)}$	1.00

flexible computer program for maximum likelihood estimation with restrictions and difficulties in determining a viable factor structure were listed as primary difficulties.

The present work is relevant to the problems defined by Campbell and Fiske (1959) insofar as it makes use of a recently developed flexible computer program by Joreskog (1967a), containing fewer objectionable restrictions. Moreover, the objectives of this dissertation research are essentially the same as Campbell's: estimation of factor loadings in a structure consistent with the type of data matrix considered. Salient differences between this study and Campbell's work are the variety and assessment of models considered and the use of maximum likelihood factor analysis.

Tucker's (1964,1965) research is also relevant. Tucker has tried to extend factor analytic models and procedures to the examination of data classified in three or more ways, "modes" of classification. This is a more general case of the usual two dimensional array (i.e., two modes) of correlations considered in factor analysis. The Campbell and Fiske paradigm is a specific case of the sort of data which can be considered using Tucker's methods: multiple methods (modes) of measuring a set of traits. The method of analysis can be considered to be a generalized variation of a principal components solution. Small characteristic roots and vectors are discarded to obtain approximations to original observations and component loadings. According to Tucker, a least squares criterion is met if all roots are used. Under some circumstances, elimination of roots may approximate a least squares solution, but a complete definition of such circumstances is lacking. The major problem with Tucker's analysis is it does not have as its basis a stochastic model. There is no basis for assessing goodness of fit of the model nor applying a variety of restrictions. Maximum likelihood estimations as implemented by Joreskog (1967a), specifies a reasonable error structure which forms the basis for a goodness of fit test.

Stanley (1961) proposed that the usual three way factorial model in analysis of variance might be appropriate for examining multitrait-multimethod data. He has shown that it is possible to express the usual mean squares obtained for the model as a function of average variances and covariances. The analysis is based partly on earlier work by Gulliksen (1950) and Stanley (1956). The derivation and algebraic results are essentially the same as those presented by Wolins (1964) and Zyzanski (1962). These authors consider averages of blocks of correlations rather than covariances. Results of F tests, using either form of data, are equivalent to tests using conventional numerical procedures. Stanley describes the interpretive aspects of the mean squares in detail, lending further support to the notion that such a three way model is viable.

Although Stanley indicates an awareness of some of the statistical problems, he provides little information on their exact nature. Some of these problems are examined in the Wolins and Zyzanski papers cited above. Specific points of investigation included comparability of scales, homogeneity of variance and the effect of differences in reliability of measures.

Investigation of various assumptions by Wolins provides evidence suggesting that the three way classificatory model is often inappropriate. Under conventional analysis, the errors are assumed to be normally and independently distributed, and the model assumed to be additive. With multitrait-multimethod data, one must also attend to the nature of the scales. That is, differences in means and variances of the observations may be a function of the particular scale used rather than a more relevant interest, the behavior upon which the scales are imposed. Wolins has suggested adjustments which can be made in order to compensate for scale differences without compromising the objectives of the analysis. The problem of nonindependence of error cannot be accommodated using the same rationale.

Specifically, under repeated measures of the same individual, by the various methods on various traits, the errors may be correlated rather than independent. This event is likely, but insofar as the correlations are homogeneous, the usual F tests are still valid. This is discussed in Boneau (1960) and Lana and Lubin (1963). If one uses average variances and covariances or correlations, the problem assumes the form of determining the nature of the structure of the variance-covariance matrix. That is, one would like to investigate the extent to which the actual covariance matrix conforms to that which would be obtained if all assumptions were met. A small sample test of deviation from such a hypothesized structure has not been developed, except for the case of a specified alternative hypothesis. Such specification is not possible with the model considered and available information. A more recent article by Wolins (1964) and a paper by Boruch and Wolins (1968) form the most relevant basis for this dissertation. The development of the model proposed in the article, and its relation to the various procedures used here, is described below, in the section on development of procedure.

A major part of the procedure for solution of the problems is based on maximum likelihood factor analysis. Joreskog (1967a,1967b) has provided detailed derivations and computer programs necessary for more flexible, unrestricted maximum likelihood estimation in factor analysis. The Joreskog approach and computer programs are used, in the current research, to implement the analysis of data. Much of Joreskog's work is based on the original development by Lawley (1940). A summary statement, together with examples, appears in Lawley and Maxwell (1963). An additional methodologically oriented development with statistical derivations appears in Morrison (1967). Both the Joreskog and Morrison presentations require some knowledge of derivatives of matrix functions. The most useful reference concerned with matrix derivatives appears to be Dwyer (1967). Descriptions by Harman (1960), Anderson and Rubin (1956) and Horst (1965) provide additional information from their respective points of view, methodological, mathematical and numerical methods. The latter two are particularly difficult to read. More detailed historical descriptions of the development of factor analysis may be found in Joreskog (1967b). The brief description of factor analysis, provided below in the section on development of a solution, is based on Rao (1955) and Joreskog (1967b).

III. DEVELOPMENT OF THE ANALYSIS

A. Factor Analysis

The approach chosen to investigate these data is based on factorial procedures. These procedures are applied to functions of the observations, the intercorrelations among the measures. Such procedures are developed from the fundamental factor postulate that the scores of all individuals on a number of tests may be expressed, in the first approximation, as a linear function of their scores in a smaller number of hypothetical measures. This is, of course, a simplistic statistical translation of what Einstein (1934) considers to be "... the grand aim of science ... to cover the greatest possible number of empirical facts by logical deduction from the smallest number of hypotheses or axioms...".

The particular variation of factorial procedures considered is maximum likelihood factor analysis. For the sake of definity, a description of the most general model and analytic procedure in factor analysis is given in this section. The presentation is based on Rao's (1955) interpretation of the problem. This particular interpretation is somewhat different from others with respect to pronounced mathematical and statistical orientation and the relative lack of methods of solution. A discussion of various methods of solution is given in Joreskog (1967b).

In factor analysis, the observed variable is hypothesized to have an underlying structure which is not observed. The unobserved hypothetical variables are generally classified into those depending on <u>common factors</u> and those depending on <u>specific factors</u>. These verbal descriptions are rendered more precise by means of a statistical model. Consider the set of observations, deviation scores, $\{y_q(q=1,2,\ldots p)\}$ that is, the set of p measures on a single individual. We have a number, say N, of such sets of observations. N is equal to the number of observational units upon which the observations are taken. Note that the subscript convention has been altered slightly from the convention previously used. The change is introduced for the sake of simplicity and to make this development more consistent with Rao's. The observation, y_q, then can be expressed in the form of the model

$$y_q = z_q + s_q \tag{3.1}$$

where z_q = unobserved variable, dependent on hypothetical common factors

s_q = unobserved variable, dependent on hypothetical specific factors

$q = 1,2,3, \ldots ,p$.

The restriction imposed on the covariance structure are:

$$\text{cov}(z_q,s_q) = 0,\ \text{cov}(z_q,s_{q'}) = 0,\ \text{cov}(s_q,s_{q'}) = 0\ (q \neq q'),$$
$$q,q' = 1, 2 \ldots p .$$

Usually the assumptions, $E(z_q) = E(s_q) = 0$, are made.

From the restrictions and the model one can deduce the following.

$$\text{Var}(y_q) = \text{Var}(z_q) + \text{Var}(s_q)$$

$$\text{Cov}(y_q,y_{q'}) = \text{Cov}(z_q,z_{q'}) \quad .$$

One can simplify the description of the covariance structure by using the following conventional matrix notation.

Σ = $p \times p$ matrix of expected covariances, so that $\text{Cov}(y_q,y_{q'})$ $= \{\Sigma\}_{qq'}$.

Γ = p x p matrix of covariances among the common factor variables such that $\mathrm{Cov}(z_q, z_{q'}) = \{\Gamma\}_{qq'}$.

ψ = p x p matrix (diagonal) of covariances among the specific factor variables, so that $s_{qq'} = \{\psi\}_{qq'}$.

In terms of the dispersion matrices then,

$$\Sigma = \Gamma + \psi \quad . \tag{3.2}$$

The object of factor analysis is to find a simple Γ and a ψ , such that one can estimate Σ . The qualification of simplicity of Γ is a rather important one. By "simplicity" is meant that the rank of Γ should be a minimum. When the rank is a minimum, say $m < p$, one is saying that the original p measures are expressible as a function of a smaller number, m , of variables.

A geometric interpretation may clarify the statements made above. Consider a vector space in which the points in the space correspond to the magnitudes of all possible linear combinations of the variables $z_1, z_2, z_3, \ldots, z_p$. The reader is again reminded that these variables are hypothetical and dependent on common factors. Let a particular combination of the variables define the vector f_r where $r = 1,2,3, \ldots ,m$. Such a vector represents a possible common factor. One can impose some restrictions and conventions on the vector space.

1. The vector product of two vectors is the covariance between them. That is, $f_r' f_{r'} = \mathrm{Cov}(f_r, f_{r'})$. The normalized vector squared is the variance of the vector, $f_r' f_r' = \mathrm{Var}(f_r)$.

2. The vectors, $f_1, f_2, \ldots, f_m$, are <u>independent</u> if no linear combination of them has a zero norm. That is, the vectors are independent if

$$\sum_{r=1}^{m} a_r f_r \neq 0, \quad \text{for all} \quad a_r \neq 0 \quad \text{simultaneously.}$$

One is, in effect, classifying a set of vectors out of the population of vectors into a category. The category contains those vectors which are independent.

3. The vector space is said to have the property of finite dimensionality, since its elements (i.e. all the f's) can be expressed as linear combinations of a finite number of elements. This finite number of elements, upon which the remaining are dependent, corresponds to the rank of the space.

4. The minimal number of elements, upon which the remaining are dependent, are necessarily independent. This set of elements is called the basis of the vector space. According to Rao (1955), the basis of a vector space is not unique, but its rank is unique. The last statement means that, given one basis, one knows the rank but there may exist other bases with exactly the same rank.

Given such a vector space, Rao states that one can always choose a basis such that the elements are orthogonal. One can also choose a basis such that some or none of the elements are orthogonal. If one is willing to assume normality, then the orthogonal basis consists of factors which are uncorrelated. The orthogonal case appears most frequently in the literature because of the relative simplicity of interpreting uncorrelated factors. The basis in which factors are not orthogonal implies that factors may be correlated. This "oblique" factor basis is less frequently used. The set of elements which constitute the oblique factor basis can be used as a system upon which to base analysis, just as the orthogonal system is.

The interpretation is usually a bit more difficult. The oblique basis can be derived from the orthogonal basis by a linear transformation. The choice of a particular basis is generally considered to be a function of the substantive information to which the factor analysis is applied. For discussion of this point the reader may find relevant information in Thurstone (1947), Harman (1960) or Rao (1955).

The most convenient manner of communicating the basis, orthogonal or oblique, is the following.

$$z_q = \lambda_{q1}Z_1 + \lambda_{q2}Z_2 + \dots + \lambda_{qm}Z_m \quad . \tag{3.3}$$

The set of vectors, $Z_1, Z_2, \dots, Z_m$, is the basis, chosen from the totality of vectors (the f's) which comprise the vector space. The reader is reminded that this totality of vectors is, in fact, the totality of all possible linear combinations of the unobserved variables, $z_1, z_2, \dots, z_p$.

The model described above, which represents a basis, is a part of the model actually used. The Z's are referred to as <u>common factor scores</u> in the literature. That is, they represent the magnitudes of the hypothetical variable which they are supposed to explain, Z_i.

The coefficients, $\lambda_{q1}, \lambda_{q2}, \dots, \lambda_{qm}$, are called <u>factor loadings</u>. In the orthogonal case they are the covariances of y_j with the factors $Z_1, Z_2, \dots, Z_k$ respectively. This equality allows one to interpret the factors more easily. Formally, then, we consider the complete model for the basic observations $\{y_q(q=1,2, \dots, p)\}$ on a single experimental unit, in matrix notation.

$$y = \Lambda z + s \ . \tag{3.4}$$

y = vector of p observations, say measures as defined above, on an experimental unit.

Λ = $p \times m$ matrix of factor loadings.

z = vector of $m < p$ common factors.

s = vector of p elements, specific factors.

The following restrictions have been imposed.

$E(z) = E(s) = 0$.

$E(zz') = \Phi$ in the oblique factor case.

$E(zz') = I$ in the orthogonal factor case.

$E(ss') = \psi$, a diagonal matrix.

The expected dispersion matrix is then,

$\Sigma = \Gamma + \psi$.

$\Sigma = \Lambda\Phi\Lambda' + \psi$ in the oblique factor case.

$\Sigma = \Lambda\Lambda' + \psi$ in the orthogonal factor case.

The preceding analytic development is related to the observed data in the following way.

Let $y_{i(jk)}$ = ith observation on the measure (jk) where j designates trait and k designates method.

$i = 1,2,3, \ldots ,N$

$j = 1,2,3, \ldots ,b$

$k = 1,2,3, \ldots ,c$

$q = 1,2,3, \ldots ,p, p = b \times c$.

The observation y_q in Rao's notation is equivalent to the observation $y_{i(jk)}$ in this notation. The i subscript in the former has simply been deleted and the subscript q is made more informative by substituting jk .

The sample mean for the measure jk may be represented as

$$y_{.(jk)} = \sum_i y_{i(jk)}/N \quad .$$

The observed covariance between measures jk and j'k' may be represented as

$$A = \underset{i}{\mathrm{Cov}}(y_{i(jk)}, y_{i(j'k')})$$

$$= \sum_i (y_{i(jk)} - y_{.(jk)})(y_{i(j'k')} - y_{.(j'k')})/(N) \quad .$$

If the observations are distributed normally and independently, the elements in the covariance follow a Wishart distribution. The relevant distribution theory is described in Anderson (1958). Lawley and Maxwell (1963) have constructed a likelihood ratio criterion for testing the hypothesis that the model is true. The criterion may be represented as

$$(N-1)\log_e[\hat{\Sigma} / A + \mathrm{tr}(A\hat{\Sigma}^{-1}) - p] \quad .$$

Lawley and Maxwell (1963) state that a satisfactory approximation to the maximization of the likelihood function is the following function, which is minimized.

$$\chi^2_\nu = (N-1)\sum_{q,q'} (A_{qq'} - \hat{\Sigma}_{qq'})^2/\hat{\psi}_{qq}\hat{\psi}_{qq'} \quad .$$

For large N, this criterion is distributed as a χ^2 with degrees of freedom equal to

$$\nu = \tfrac{1}{2}p(p+1) - pm - \tfrac{1}{2}m(m+1) - p - n_\psi - \sum_{r=1}^{m}\max(n_r,m) \quad ,$$

where n_ψ is the number of fixed parameters in ψ, n_r the number of independent restrictions on factor r .

Detailed discussions can be found in Joreskog (1967a) and Morrison (1967). Some additional notes are described by Bartlett (1951).

B. The Classificatory Model As A Basis For Analysis

Consider the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \ell_{ijk} \qquad (3.5)$$

where

$$i = 1,2,3, \ldots ,N \qquad \ell_{ijk} \sim \mathrm{NID}(0,\sigma^2_\varepsilon)$$

$$j = 1,2,3, \ldots ,b \qquad \alpha_i \sim \mathrm{NID}(0,\sigma^2_\alpha)$$

$$k = 1,2,3, \ldots ,c \qquad [(\alpha\beta)_{ij} \sim \mathrm{NID}(0,\sigma^2_{\alpha\beta})]$$

$$[(\alpha\gamma)_{ik} \sim \mathrm{NID}(0,\sigma^2_{\alpha\gamma})] \quad .$$

The α_i represents block or person effect and is considered random. The γ_k , β_j can be considered treatment effects and are fixed factors, "methods" and "traits". This is the sort of model espoused explicitly by Stanley (1961) and implicitly by Campbell and Fiske (1959) for the type of data we are considering. Further research using multimethod-multitrait data was done by Wolins (1964). This presentation draws also from work by Joreskog (1967b) and, of course, on the general theory of experimental design as presented by Kempthorne. Definitions of the methods and traits, and explanations of the problem are given above in Section II of this dissertation.

Those sources of variance involving the random variable, α , can be assessed by obtaining a covariance matrix of dimensions <u>bc</u> by <u>bc</u> . The expectations of these covariances according to the above model are derived in the following manner.

The expectation of the covariance over i , for a given jk , j'k' combination, based on the model, is:

$$\underset{i}{\mathrm{Cov}}(Y_{ijk}, Y_{ij'k'}) = E\{(Y_{ijk} - E(Y_{.jk}))(Y_{ij'k'} - E(Y_{.j'k'}))\} \quad . \qquad (3.6)$$

We know that

$$E(Y_{.jk}) = E(\sum_i Y_{ijk}/N)$$

$$= E(\mu + \sum_i \alpha_i/N + \beta_j + \sum_i (\alpha\beta)_{ij}/N + \gamma_k$$

$$+ \sum_i (\alpha\gamma)_{ik}/N + (\beta\gamma)_{jk} + \sum_i \epsilon_{ijk}/N) \quad .$$

Using the usual assumptions attached to the model one can say

$$E(\alpha_i) = 0$$

$$E(\beta_j) = \beta_j$$

$$E(\gamma_k) = \gamma_k$$

$$E((\beta\gamma)_{jk}) = (\beta\gamma)_{jk}$$

$$E(\sum_i \epsilon_{ijk}) = 0 \quad .$$

In addition, if the number of blocks, N , is large then

$$E(\sum_i (\alpha\beta)_{ij}) = 0$$

$$E(\sum_i (\alpha\gamma)_{ik}) = 0 \quad .$$

So that

$$E(Y_{.jk}) = (\mu + \beta_j + \gamma_k + (\beta\gamma)_{jk}) \quad ,$$

$$E\{(Y_{ijk} - E(Y_{.jk}))(Y_{ij'k'} - E(Y_{.j'k'}))\}$$

$$= E\{(\alpha_i + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \varepsilon_{ijk})(\alpha_i + (\alpha\beta)_{ij'} + (\alpha\gamma)_{ij'} + \varepsilon_{ij'k'})\}$$

$$= E(\alpha_i^2 + \alpha_i(\alpha\beta)_{ij'} + \alpha_i(\alpha\gamma)_{ik'} + \alpha_i\varepsilon_{ij'k'}$$

$$+ (\alpha\beta)_{ij}\alpha_i + (\alpha\beta)_{ij}(\alpha\beta)_{ij'} + (\alpha\beta)_{ij}(\alpha\gamma)_{ik'} + (\alpha\beta)_{ij}\varepsilon_{ij'k'}$$

$$+ (\alpha\gamma)_{ik}\alpha_i + (\alpha\gamma)_{ik}(\alpha\beta)_{ij'} + (\alpha\gamma)_{ik}(\alpha\gamma)_{ik'} + (\alpha\gamma)_{ik}\varepsilon_{ij'k'}$$

$$+ \varepsilon_{ijk}\alpha_i + \varepsilon_{ijk}(\alpha\beta)_{ij'} + \varepsilon_{ijk}(\alpha\gamma)_{ik'} + \varepsilon_{ijk}\varepsilon_{ij'k'}) \quad .$$

Using the assumptions about independence and variance of effects, interactions and error we have

$$\mathrm{Cov}(Y_{ijk}, Y_{ij'k'})_i = \sigma_\alpha^2 + \delta_{jj'}\sigma_{\alpha\beta}^2 + \delta_{kk'}\sigma_{\alpha\gamma}^2 + \delta_{jj'}\delta_{kk'}\sigma_\varepsilon^2 \quad . \tag{3.7}$$

The symbols, $\delta_{jj'}$ and $\delta_{kk'}$, above, are Kronecker deltas. That is, $\delta_{jj'} = 1$ if $j = j'$ and $\delta_{jj'} = 0$ otherwise, and similarly for $\delta_{kk'}$. This result is displayed in Table 2, the covariance matrix.

The entries in Table 2 are the following identities, computed in the manner described above. The symbol ρ is used for convenience.

$$E\{\underset{i}{\mathrm{Cov}}(Y_{ijk}, Y_{ij'k'})\} = \sigma_\alpha^2 = \rho_4$$

$$E\{\underset{i}{\mathrm{Cov}}(Y_{ijk}, Y_{ij'k})\} = \sigma_\alpha^2 + \sigma_{\alpha\gamma}^2 = \rho_3$$

$$E\{\underset{i}{\mathrm{Cov}}(Y_{ijk}, Y_{ijk'})\} = \sigma_\alpha^2 + \sigma_{\alpha\beta}^2 = \rho_2$$

$$E\{\underset{i}{\mathrm{Cov}}(Y_{ijk}, Y_{ijk})\} = \sigma_\alpha^2 + \sigma_{\alpha\beta}^2 + \sigma_\varepsilon^2 = \rho_1 \quad .$$

Table 2. Covariance matrix

	$\beta_1\gamma_1$	$\beta_1\gamma_2$	$\beta_1\gamma_3$	$\cdots$	$\beta_1\gamma_c$	$\beta_2\gamma_1$	$\beta_2\gamma_2$	$\beta_2\gamma_3$	$\cdots$	$\beta_2\gamma_c$	$\cdots$	$\beta_b\gamma_c$
$\beta_1\gamma_1$	ρ_1	ρ_2	ρ_2	$\cdots$	ρ_2	ρ_3	ρ_4	ρ_4	$\cdots$	ρ_4	$\cdots$	ρ_4
$\beta_1\gamma_2$	ρ_2	ρ_1	ρ_2	$\cdots$	ρ_2	ρ_4	ρ_3	ρ_4	$\cdots$	ρ_4	$\cdots$	ρ_4
$\beta_1\gamma_3$	ρ_2	ρ_2	ρ_1	$\cdots$	ρ_2	ρ_4	ρ_4	ρ_3	$\cdots$	ρ_4	$\cdots$	ρ_4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
$\beta_1\gamma_c$	ρ_2	ρ_2	ρ_2	$\cdots$	ρ_1	ρ_4	ρ_4	ρ_4	$\cdots$	ρ_3	$\cdots$	ρ_3
$\beta_2\gamma_1$	ρ_3	ρ_4	ρ_4	$\cdots$	ρ_4	ρ_1	ρ_2	ρ_2	$\cdots$	ρ_2	$\cdots$	ρ_4
$\beta_2\gamma_2$	ρ_4	ρ_3	ρ_4	$\cdots$	ρ_4	ρ_2	ρ_1	ρ_2	$\cdots$	ρ_2	$\cdots$	ρ_4
$\beta_2\gamma_3$	ρ_4	ρ_4	ρ_3	$\cdots$	ρ_4	ρ_2	ρ_2	ρ_1	$\cdots$	ρ_2	$\cdots$	ρ_4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
$\beta_2\gamma_c$	ρ_4	ρ_4	ρ_4	$\cdots$	ρ_3	ρ_2	ρ_2	ρ_2	$\cdots$	ρ_1	$\cdots$	ρ_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$
$\beta_b\gamma_c$	ρ_4	ρ_4	ρ_4	$\cdots$	ρ_3	ρ_4	ρ_4	ρ_4	$\cdots$	ρ_3	$\cdots$	ρ_1

The expected mean squares and variance components for those sources of variance involving the random variable are indicated below.

Source	Variance component	EMS
α	$\rho_4 = \sigma^2_\alpha$	$\sigma^2_\varepsilon + bc\sigma^2_{\underline{\alpha}}$
$\alpha\beta$	$\rho_2 - \rho_4 = \sigma^2_{\alpha\beta}$	$\sigma^2_\varepsilon + c\sigma^2_{\underline{\alpha\beta}}$
$\alpha\gamma$	$\rho_3 - \rho_4 = \sigma^2_{\alpha\gamma}$	$\sigma^2_3 + b\sigma^2_{\underline{\alpha\gamma}}$
error	$\rho_1 - \rho_2 - \rho_3 + \rho_4 = \sigma^2_\varepsilon$	σ^2_ε .

C. Random Block Effect In the Model

We have assumed that one has a large number of blocks (i.e., persons). This situation is contrary to ordinary circumstances insofar as large numbers of blocks are not usually available for experimental data. Given that the blocks are available, one can hope to obtain meaningful estimates of the variance components. By "meaningful" is meant that the conditions required for accuracy and precision in estimation of the components should be met. Specifically, a large number of blocks would make the assumption of normality of the (average) effect more viable. It would also decrease the variance associated with the variance component estimates. If one obtained 100 or more blocks, under the design implied by the model above, one could, with confidence, estimate variance components σ^2_α , $\sigma^2_{\alpha\beta}$, $\sigma^2_{\alpha\gamma}$ and σ^2_ε . For the moment, we consider the magnitude of the components, rather than the usual significance tests, to be of primary interest. The relevance of special significance tests is discussed in the section of this paper under the general topic of goodness of fit.

If, then, one is primarily interested in the random block effect, one can consider putting aside the other fixed factors in the model. That is, the following model results if one adjust for the fixed sources of variance:

$$Y'_{ijk} = \alpha_i + \alpha\beta_{jk} + \alpha\gamma_{jk} + \epsilon_{ijk} \quad . \tag{3.8}$$

Note that the Y_{ijk} is primed to remind the reader that this is a deviation score.

D. Elimination of Fixed Effects

The β_j , γ_k , $\gamma\beta_{jk}$ effects and interaction may be eliminated from the original model under the following reasoning.

1. The average fixed "treatment" effects in the context of the data are essentially the effects of particular scales used in a method-trait combination being considered. The psychological scales are arbitrary, of course, to the extent that one can adjust all means for various combinations. That is, one can equate all means to some (arbitrary) constant value, effectively eliminating these factors form the data. The interactions of random and fixed effects are unaffected by the procedure. The notions developed here are also discussed in Wolins (1964) and Stanley (1961). The response Y'_{ijk} now is a deviation score adjusted for fixed effects and interaction.

2. Conditional on the model, which does not recognize the three factor interaction, the total variance of the blocks on each measure, method-trait combination, is expected to be the same. That is, the diagonals of the covariance matrix are expected to be equal. If these diagonals were constrained so that they must be equal, the residuals would not be inde-

pendent. They are correlated. However, the maximum likelihood procedure does not so constrain these diagonal entries in the covariance matrix. Rather, each diagonal entry represents an equation and residuals from unity occur in the diagonal just as residuals from expected covariances occur in the remaining part of the matrix.

3. One can show that adjustment of the observed values, Y_{ijk} to Y'_{ijk}, does not affect the covariance between Y'_{ijk} and $Y'_{ij'k'}$. That is,

$$\rho_{Y_{ijk}, Y_{ij'k'}} = \rho_{Y'_{ijk}, Y'_{ij'k}}$$

$$\rho_{Y_{ijk}, Y_{ij'k}} = \rho_{Y'_{ijk} Y'_{ij'k}}$$

$$\rho_{Y_{ijk}, Y_{ijk'}} = \rho_{Y'_{ijk}, Y'_{ijk'}} .$$

Further proofs are shown in Stanley (1961).

E. Generalization of the Model

Suppose we now consider any information which suggests the model, $Y'_{ijk} = \alpha_i + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \varepsilon_{ijk}$, is inappropriate in its current form.

1. It appears that the result Y'_{ijk} may not be a simple linear function of the effects. Specifically, it is reasonable to conjecture that α_i, $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$ are weighted in a manner determined by the specific method-trait (jk subscript) combination used. "Reasonableness" is based on research by Stanley (1961) and Campbell (1965) and Wolins as well as on the efforts of the writer. A more appropriate model then might be

$$Y'_{ijk} = A_{jk}\alpha_i + B_{jk}\alpha\beta_{ij} + C_{jk}\alpha\gamma_{ik} + \varepsilon_{ijk} . \quad (3.9)$$

The model may be interpreted in the following manner. For a given measure, a specific trait-method combination, one is assuming that the observation is a function of the variables indicated. The X_i is dependent on subjects. The weighting is indicated by the magnitude of A_{jk}. The person x trait interaction $(\alpha\beta)_{ij}$ is also weighted. That is, the interaction contributes to the observed score in some manner, depending on the measure. The relative importance of the contribution is indicated by the magnitude of B_{jk}. Similar comments apply to the person x method interaction $(\alpha\gamma)_{ik}$ and C_{jk}. The error, or deviation from expected values, is dependent on person, method and trait, as in the original model.

The statistical attributes of the model are considered in the following paragraphs.

2. Based on prior information, one may also conjecture that some of the effects in the model are not independent of one another. That is, it may be reasonable to conjecture that the interaction $(\alpha\beta)_{ij}$ is correlated with the interaction $(\alpha\gamma)_{ik}$, for example. That is,

$$\rho^{BC}_{jk} = \text{expected correlation between } (\alpha\beta)_{ij} \text{ and } (\alpha\gamma)_{ik}$$

$$= \sum_i \frac{(\alpha\beta)_{ij}(\alpha\gamma)_{ik}}{N\sigma_{\alpha\beta}\sigma_{\alpha\gamma}} \neq 0 \ .$$

In the context of the data being examined, this means that the person x method interaction may be correlated with the person x trait interaction. That is, whatever makes the contributions of method dependent on person may be linearly related to the phenomena of contributions of trait being dependent on person. Similarly it seems reasonable to suppose traits are correlated and methods are correlated. For example, peer ratings, a method, may be correlated with sociability, a trait.

It should be noted that, without further constraints, the parameters in this last model are not estimable. However, it is shown below that imposing various constraints results in models which estimate the usual analysis of variance components and which are related to models which estimate ordinary factor loadings.

Formally, we then consider the model,

$$Y_{ijk} = A_{jk}\alpha_i + B_{jk}(\alpha\beta)_{ij} + C_{jk}(\alpha\gamma)_{ik} + \varepsilon_{ijk} \qquad (3.10)$$

A_{jk} , B_{jk} , C_{jk} are parameters

where

$$\varepsilon_{ijk} \sim NID(0,\sigma^2_{\varepsilon_{jk}})$$

$$\alpha_i \sim NID(0,1)$$

$$(\alpha\beta)_{ij} , (\alpha\gamma)_{ik} \sim N(0,1) .$$

The relaxation of the usual assumption of independence of $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ implies that $(\alpha\beta)_{ij'}$ may be correlated with $(\alpha\beta)_{ij}$, or $(\alpha\gamma)_{ik}$.

Given this last model, let $\alpha_i = x_i$, $(\alpha\beta)_{ij} = x_{ij}$, $(\alpha\gamma)_{ik} = x_{ik}$. The equalities and restrictions are put onto this form in order to make the relationship between this last model and the factor analytic model more obvious. That is,

$$Y_{i(jk)} = A_{jk}X_i + B_{jk}X_{ij} + C_{jk}X_{ik} + \varepsilon_{i(jk)} . \qquad (3.11)$$

We may express any covariance over i as follows:

$$\rho_{jk,j'k'} = E\{(Y_{ijk} - E(Y_{.jk}))(Y_{ij'k'} - E(Y_{.j'k'}))\} .$$

$$Y_{.jk} = \sum_i (A_{jk}\alpha_i + B_{jk}(\alpha\beta)_{ij} + C_{jk}(\alpha\gamma)_{ik} + \epsilon_{ijk})/N$$

$$= (A_{jk} \sum_i \alpha_i + B_{jk} \sum_i (\alpha\beta)_{ij} + C_{jk} \sum_i (\alpha\gamma)_{ik} + \sum_i \epsilon_{ijk})/N$$

$= 0$, if the number of blocks is large.

$$\rho_{jk,j'k'} = E\{(Y_{ijk})(Y_{ij'k'})\}$$

$$= E\{(A_{jk}\alpha_i + B_{jk}(\alpha\beta)_{ij} + C_{jk}(\alpha\gamma)_{ik} + \epsilon_{ijk})(A_{j'k'}\alpha_i$$

$$+ B_{j'k'}(\alpha\beta)_{ij'} + C_{j'k'}(\alpha\gamma)_{ik'} + \epsilon_{ij'k'})\}$$

$$= E\{(A_{jk}A_{j'k'}\alpha_i^2 + A_{jk}B_{j'k'}\alpha_i(\alpha\beta)_{ij'} + A_{jk}C_{j'k'}\alpha_i(\alpha\gamma)_{ik'}$$

$$+ A_{jk}\alpha_i\epsilon_{ij'k'}) + (B_{jk}A_{j'k'}\alpha_i(\alpha\beta)_{ij} + B_{jk}B_{j'k'}(\alpha\beta)_{ij}(\alpha\beta)_{ij'}$$

$$+ B_{jk}C_{j'j'}(\alpha\beta)_{ij}(\alpha\gamma)_{ik'} + B_{jk}(\alpha\beta)_{ij}\epsilon_{ij'k'}) + (C_{jk}A_{j'k'}\alpha_i(\alpha\gamma)_{ik}$$

$$+ C_{jk}B_{j'j'}(\alpha\beta)_{ij'}(\alpha\gamma)_{ik} + C_{jk}C_{j'k'}(\alpha\gamma)_{ik}(\alpha\gamma)_{ik'}$$

$$+ C_{jk}(\alpha\gamma)_{ik}\epsilon_{ij'k'}) + (\epsilon_{ijk}A_{j'k'}\alpha_i + \epsilon_{ijk}B_{j'j'}(\alpha\beta)_{ij'}$$

$$+ \epsilon_{ijk}C_{j'k'}(\alpha\gamma)_{ik'} + \epsilon_{ijk}\epsilon_{ij'k'})\} \quad .$$

Using the restrictions which are imposed on the model, one obtains the following.

$$\rho_{jk,j'k'} = A_{jk}A_{j'k'} + B_{jk}B_{j'k'}\rho_{jj'}^{BB} + B_{jk}C_{j'k'}\rho_{jk'}^{BC}$$
$$+ B_{j'k'}C_{jk}\rho_{j'k'}^{CB} + C_{jk}C_{j'k'}\rho_{kk'}^{CC} + \rho_{e_{jk}e_{j'k'}} \tag{3.12}$$

where $\rho_{jj'}^{BB}$ is, for example, the correlation between $(\alpha\beta)_{ij}$ and $(\alpha\beta)_{ij'}$.

Putting aside the model involving the observational units and attending to the fact that we have developed a model where the observed covariance is the dependent variable, we note we have $bc(bc+1)/2$ equations in some number of unknowns. The number of unknowns can be better assessed if we define three matrices for traits $j = 1,2,3,4,5$ and methods $k = I,II,III$.

$$
\Lambda_{15x9} = \begin{bmatrix}
A_{1I} & B_{1I} & 0 & 0 & 0 & 0 & C_{1I} & 0 & 0 \\
A_{2I} & 0 & B_{2I} & 0 & 0 & 0 & C_{2I} & 0 & 0 \\
A_{3I} & 0 & 0 & B_{3I} & 0 & 0 & C_{3I} & 0 & 0 \\
A_{4I} & 0 & 0 & 0 & B_{4I} & 0 & C_{4I} & 0 & 0 \\
A_{5I} & 0 & 0 & 0 & 0 & B_{5I} & C_{5I} & 0 & 0 \\
A_{1II} & B_{1II} & 0 & 0 & 0 & 0 & 0 & C_{1II} & 0 \\
A_{2II} & 0 & B_{2II} & 0 & 0 & 0 & 0 & C_{2II} & 0 \\
A_{3II} & 0 & 0 & B_{3II} & 0 & 0 & 0 & C_{3II} & 0 \\
A_{4II} & 0 & 0 & 0 & B_{4II} & 0 & 0 & C_{4II} & 0 \\
A_{5II} & 0 & 0 & 0 & 0 & B_{5II} & 0 & C_{5II} & 0 \\
A_{1III} & B_{1III} & 0 & 0 & 0 & 0 & 0 & 0 & C_{1III} \\
A_{2III} & 0 & B_{2III} & 0 & 0 & 0 & 0 & 0 & C_{2III} \\
A_{3III} & 0 & 0 & B_{3III} & 0 & 0 & 0 & 0 & C_{3III} \\
A_{4III} & 0 & 0 & 0 & B_{4III} & 0 & 0 & 0 & C_{4III} \\
A_{5III} & 0 & 0 & 0 & 0 & B_{5III} & 0 & 0 & C_{5III}
\end{bmatrix}
$$

The superscripts in $\underset{\sim}{\sigma}$, below, have been deleted for convenience.

$$\bar{\Phi} = \begin{bmatrix} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ 0 & \rho_{1,2} & 1 & & & \text{(symmetrical)} & & & \\ 0 & \rho_{1,3} & \rho_{2,3} & 1 & & & & & \\ 0 & \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 & & & & \\ 0 & \rho_{1,5} & \rho_{2,5} & \rho_{3,5} & \rho_{4,5} & 1 & & & \\ 0 & \rho_{1,I} & \rho_{2,I} & \rho_{3,I} & \rho_{4,I} & \rho_{5,I} & 1 & & \\ 0 & \rho_{1,II} & \rho_{2,II} & \rho_{3,II} & \rho_{4,II} & \rho_{5,II} & \rho_{I,II} & 1 & \\ 0 & \rho_{1,III} & \rho_{2,III} & \rho_{3,III} & \rho_{4,III} & \rho_{5,III} & \rho_{I,III} & \rho_{II,III} & 1 \end{bmatrix} . \tag{3.14}$$

ψ_{15x15} = Diagonal matrix with non-zero entries of $\sigma^2_{\epsilon_{jk}}$ (or $\rho_{e_{jk}e_{jk}}$) .

It is apparent that there are 3 bc unknowns in the Λ matrix, $(b+c)(b+c-1)/2$ unknowns in the $\bar{\Phi}$ matrix and bc unknowns in the ψ matrix. It also can be shown that the expectation of the covariance matrix is,

$$\Sigma = E(\rho_{15x15}) = \Lambda_{15x9}\bar{\Phi}_{9x9}\Lambda'_{9x15} + \psi_{15x15} \quad . \tag{3.15}$$

This, of course, is of the form of the usual factor analytic model dispersion matrix described in Lawley and Maxwell, Joreskog, etc.

The reader is reminded that if the usual analysis of variance constraints were applied then

$$A_{jk} = \sigma_{\alpha}$$
$$B_{jk} = \sigma_{\alpha\beta}$$
$$C_{jk} = \sigma_{\alpha\gamma} \quad ,$$

for all j, k .

Also Φ would be an identity matrix and ψ would be a diagonal matrix with the non-zero elements equal to σ_ε^2 .

In summary, the classificatory model recognizes the same random sources of variance as the constrained factor analytic model. However, the factor analytic model does not require the variance components associated with the random variables to be homogeneous across measures. Instead, the latter model recognizes that the variance components may differ from measure to measure.

F. Estimates of Variance Components

The reader is reminded that, if the analysis of variance and usual assumptions are true, then the following equalities hold.

$$A_{jk} = \sigma_\alpha$$

$$B_{jk} = \sigma_{\alpha\beta}$$

$$C_{jk} = \sigma_{\alpha\gamma}$$

$$\sigma_{\varepsilon_{jk}} = \sigma_\varepsilon \quad \text{for} \quad j = 1,2, \ldots ,b \ , \ k = 1,2, \ldots ,c \ .$$

That is, the factor loadings for each measure would be equal to the estimate of the square root of the corresponding variance components. For large N , one can show that an unbiased estimate of the standard deviation is, in fact, the square root of the variance. See, for example, Cureton (1968).

For the type of data which is being considered here, the A's, B's and C's are frequently unequal for the treatment combinations. That is, the variance components corresponding to the effects and interactions of the treatment combinations are not homogeneous.

Consider the variance of a single measure (jk). From Equation 3.12, we have the following.

$$\rho_{jk,jk} = \mathrm{Var}(Y_{i(jk)}) = A_{jk}^2 + B_{jk}^2 + C_{jk}^2 + 2B_{jk}C_{jk}\rho_{jk}^{BC} + \sigma_{\epsilon jk}^2 \quad .$$

That is, the $\hat{\Lambda}$ is comprised of $\hat{A}_{jk}$, $\hat{B}_{jk}$ and $\hat{C}_{jk}$. The $\hat{\Phi}$ is comprised of ρ_{jk}^{BC} as well as $\hat{\rho}_{jj'}^{BB}$, and $\hat{\rho}_{kk'}^{CC}$. The diagonal elements of $\hat{\psi}$ are the estimates of error associated with each measure, $\sigma_{\epsilon_{jk}}^2$.

By substituting the appropriate estimates into the equation above, one obtains the total variance accounted for by the factor analytic model.

The variance attributable to the first factor, the general factor represented by A_{jk} , is A_{jk}^2 . The variance attributable to method and trait combined is

$$B_{jk}^2 + C_{jk}^2 + 2B_{jk}C_{jk}\sigma_{jk}^{BC} \quad .$$

It appears reasonable to consider methods and traits separately, for the sake of simplicity of interpretation.

If trait and method factors were uncorrelated, i.e., $\rho_{jk}^{BC} = 0$ then B_{jk}^2 would be the variance attributable to trait and C_{jk}^2 is equal to the variance attributable to method. For the data sets examined here, however, methods and traits are correlated. This poses a problem insofar as a method-trait correlation is not easily interpretable. One is saying, in effect, that the method and trait are linearly related although common factor (i.e., the general factor) variance has been considered. One is confronted with the problem of explaining or conjecturing the reason for such a correlation. In the absence of other substantive knowledge about the traits and methods a reasonable approach appears to be the following.

One can consider half of the covariance to be attributable to the trait and the other half to be attributable to the method. That is, the variances attributable to traits and methods respectively, are

$$B_{jk}^2 + B_{jk}C_{jk}\rho_{jk}^{BC}$$

$$C_{jk}^2 + B_{jk}C_{jk}\rho_{jk}^{BC} \quad .$$

In many cases, the correlations between methods and traits are so small that consideration of the covariance may be fatuous. In some cases, enough information may be known to assign the covariance to either trait or method variance terms. A possible approach is given below.

One can implement this procedure for each of the $b \times c = p$ sets of measures. That is, for each row of the $\hat{\Lambda}$ matrix, using the appropriate elements in $\hat{\phi}$ and $\hat{\psi}$, one computes the percentage variance attributable to the sources, the hypothetical factors.

If the $\hat{\phi}$ matrix were of sufficient size one might factor analysize the $\hat{\phi}$ matrix. The results of an orthogonal factor analysis of the $\hat{\phi}$ matrix would provide a basis for decomposing the covariance among the hypothetical variables on an analytical basis.

G. Implementation of A Solution

The solution for restricted maximum likelihood factor analysis (RMLFA) and a computer program for implementation of the analysis have been developed by Joreskog (1967a,1967b). Joreskog's RMLFA Computer Program (1967a), allows one to constrain elements in the Λ , ϕ and ψ matrices.

That is, elements in these matrices may be specified in advance to be some particular values. For the development described in this dissertation, certain values in Λ and $\bar{\phi}$ are limited to be equal to zero. The parameters which are left unspecified are free to vary and are estimated, conditional on these zero constraints, using the maximum likelihood method.

In using the program, one represents those variables which are free to vary as 1's in the Λ and $\bar{\phi}$ matrices. The remaining elements are restricted to be equal to zero. For example, if three factors are extracted from a 9×9 correlation matrix, and these factors are oblique, one may wish to restrict each factor to load on only three variables. A <u>factor pattern</u> required by the program which conforms to these requirements is the following.

$$\Lambda_P = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \bar{\phi}_P = \begin{bmatrix} (\ ) & 1 & 1 \\ 1 & (\ ) & 1 \\ 1 & 1 & (\ ) \end{bmatrix} .$$

Values for the diagonal upper triangular elements in the $\bar{\phi}$ matrix are specified by the algorithm used in the computer program. If the model is considered inappropriate, then one can eliminate factors and/or collapse one factor into another. The procedure is described and applied to sample data in the next section. The program is quite flexible insofar as it allows one to manipulate and alter models easily, as further examples may suggest.

The program provides technical information which is useful in examining solutions of a given model. The information is of the type to be described below in this dissertation under the topic of rationale for changes in models. Specifically, one obtains information on boundary problems, approximations to the variance of the estimates of the parameters and indicator of the shape of the likelihood function. That is, the program computes first order and second order partial derivatives of the function at the points which represent estimates of the parameters. A large sample χ^2 statistic, indicative of goodness of fit, is also computed.

The program uses the Fletcher and Powell (1963) method of optimization to maximize the function. Further information on this and other aspects of the program are furnished in the references cited.

1. Changes in the RMLFA Computer Program

Some minor changes in the computer program were necessary for operation. The operating system and compiler for the IBM 7044, for which the program was written (in FORTRAN IV), is somewhat different from the system and compiler in the available computer, an IBM 360-65.

Input and output/unit designations were changed. These changes can be made by the machine itself using appropriate control cards.

Some format and dimension statements required changes, specifically those referring to variable format, alphanumeric characters. The change is required because the 7044 uses a word length of 6 bytes, while the IBM 360 uses a word length of 4 bytes for data storage.

The only other minor changes consisted of deletion of two machine language programs, TODAY and CLOCK, whose functions are redundant under the current system.

The program restricts the maximum size of the correlation matrix and the Λ , Φ and ψ matrices. Work is currently being done to increase the size of matrices and the number of factors which the program will accommodate. The program uses approximately 110K (words) of storage in the IBM 360-65.

2. RMLFA Program testing

Three sets of data were used in assessing the operation of the computer program. The first two sets were provided by Joreskog (1967a) and computed results appear to be consistent with the correct solutions derived by others. A third set of data was generated specifically for examination of the operation of the program with respect to the development presented in this dissertation.

H. Rationale For Changes In the Model

This section of the dissertation contains an explanation of the general rationale used as a basis for alteration of the factor model.

The initial model hypothesized to explain a given multitrait-multimethod data set is represented by Equation 3.12 with restrictions imposed in the manner illustrated in Equations 3.13, 3.14 and 3.15. The relevant assumptions and model are given in the preceding section on development of a solution. The characteristics of the solution, conditional on this initial model, may not be satisfactory. If such attributes are unsatisfactory in some sense, then the model is considered inappropriate. A new model is then hypothesized. The new model is of the same general form as the initial model and incorporates changes to eliminate undesirable features of the initial solution. The procedure is repeated for each consecutive model until a viable solution is achieved. To this extent, any one model is conditional on all those which preceded it.

Rejection or alteration of a model is based on several criteria. Each of the criteria is based on the properties of the solution to the particular factor analytic model hypothesized. The criteria fall into the general categories of boundary problems, inconsistency of solution and goodness of fit of the model. These are described in detail below.

The type of alteration made on a given model is a function of the type of criterion which the solution may violate. The relationship between the type of change necessary and the particular violation are also described below.

1. Boundary conditions

The general nature of boundary problems and a detailed examination of aspects of these problems is presented in this section. The boundaries refer to the range of values over which estimates of the parameters Λ , Φ and ψ may vary. Specifically, one can deduce from the model that the following inequalities must be true when the data are comprised of correlations.

$$-1.0 \leq \{\hat{\Lambda}\}_{jk} \leq 1.0 \ , \ -1.0 \leq \{\hat{\Phi}\}_{kk'} \leq 1.0 \ , \ 0 < \{\hat{\psi}\}_{jj'}$$

$$\text{for } j,j' = 1,2, \ldots p \text{ and } k,k' = 1,2, \ldots m \ .$$

Joreskog (1967a) designates a solution which falls into the interior of the allowable regions as a proper solution. An improper solution refers to one in which one or more of the estimates of the parameters fall on the boundaries of the allowable region. One may choose to accept or reject an improper solution, depending on other attributes of the solution. That is, an improper solution is viable in the sense that the parameters conform to requirements stated above. However, the derivatives at the solution are not all zero. Therefore, the solution is not a maximum likelihood solution. Boundary problems occurred with high frequency in solutions of data analyzed for the dissertation. In the 40 models devised for three data sets, boundary problems with respect to $\hat{\Lambda}$, $\hat{\Phi}$ or $\hat{\psi}$ were evident in 30 solutions. Joreskog (1967b) has also found high frequencies of occurrence for data which he has analyzed.

If one or more of the estimates of parameters fall completely outside the acceptable region, the solution is not considered a viable one. The model from which the estimates were generated is rejected or altered. The occurrence of estimates whose magnitudes are outside allowable values is dependent, of course, on the data and is also dependent on the algorithm used to compute the solution. That is, algorithms differ with respect to whether they consider values outside the restricted region. The algorithm used in this dissertation, based on Joreskog's RMLFA Computer Program (1967a), imposes restrictions stated above on $\hat{\psi}$ (and effectively on $\hat{\Phi}$) only. Therefore, it can occur that the absolute value of some $\hat{\Lambda}_{jk}{}'$ is greater than 1.0.

Consider the restrictions imposed on magnitudes of $\hat{\Lambda}$. If the absolute magnitudes of one or more $\hat{\Lambda}_{jk}$ values are well above 1.0, then the solution violates the stated requirements. The likelihood function, conditional on the model, does not appear to have a well defined maximum within the allowable region. Such events were rare in this research, occuring only twice in more than forty solutions to various sets of data. In each of the cases examined, the occurrence of such values is attributable to a paucity of factors in the model. That is, more factors need to be hypothesized in order to account for the data. These comments are particularly true for the orthogonal unrestricted factor analyses, which comprised some of the intermediate solutions. The addition of more factors to the model is limited in some respects. The additional factor must make some sense with respect to relationships with other factors, with respect to the data and with respect to the rationale described above for development of the models. For example, in an intermediate solution three methods factors and five trait factors were hypothesized in the model. This corresponds to a $\hat{\Lambda}$ matrix identical to the one illustrated in Equation 3.13, except that the first column in the latter has been eliminated. An intermediate solution resulted in a single $\hat{\Lambda}_{jk}$ element greater than 1.2. The addition of a general factor, corresponding to the first column in $\hat{\Lambda}$ in Equation 3.13, resulted in a solution which did not violate the requirements.

If the absolute magnitude of $\hat{\Lambda}_{jk}$ does not greatly exceed 1.0, then one might consider acceptance of the solution, attributing the discrepancy to the inaccuracy of the computer and program or an unusual event regarding the sample. The preceding statement is admittedly rather arbitrary, but not enough information is known about the computer-algorithm relationship to make a definitive statement. In only one case did a solution to the data considered here fall into this category. The magnitudes of a single element within the $\hat{\Lambda}$ matrix was of the order 1.015. The appearance of the value was accompanied by other indications that the model should be rejected. The models were indeed rejected and no decision, based entirely on the value of 1.015, was necessary.

Consider the restriction on $\hat{\psi}$. The elements of $\hat{\psi}$ must be greater than zero because computation of the estimates requires inversion of these elements. If a $\hat{\psi}$ element is near zero, then no useful solution is possible. That is, the variance of a measure must not be accounted for by common factors only. The variance of each measure must be a function of common factor variance and random deviations in order to obtain a solution. In computation of the parameters, the RMLFA Program restricts $\hat{\psi}$ to be $\hat{\psi} \geq .05$. If the value .05 is estimated, the program terminates the maximization procedure and provides an appropriate error message (IND = 2) . One can reject or accept the model depending on other attributes of the solution. For example, if no other boundary problems are apparent and one can judge that the model fits the data well, then one might accept the solution. The goodness of fit of the model is considered below.

Consider the restriction on Φ , the matrix of correlations between the factors. It is possible for the likelihood function to be a maximum when one or more of the elements in $\hat{\Phi}$ are close to 1.0 or -1.0, i.e., the boundaries. The RMLFA Computer Program ceases the optimization procedure when the value of a $\hat{\Phi}_{jk}$ approaches -.975 or .975. The procedure is terminated in order to provide a maximum likelihood solution, yet maintain the general form of the model specified. If the absolute magnitudes of one or more of the $\hat{\Phi}_{jk}$ are rather high, say .975, then one can infer that the factors so correlated are equivalent. They are equivalent in the sense that there is a high linear relation between them. Since the scales associated with the factor scores are the same, one can infer further that the two factors are identical. Under these circumstances, one might wish to obtain a more parsimonious description of the data in the sense that fewer factors are hypothesized.

One must revise the model so that a single factor replaces the two which are known to be identical. Operationally, one collapses a factor into the factor to which it is related. The following example illustrates the collapsing procedure in terms of the pattern matrices required by the RMLFA Computer Program to define the model. The procedure to be described was actually applied to Data Set 3. The basis for such application is illustrated in Tables 7 and 8. Tables 9 and 10 contain estimates of the parameters after collapse of the factors indicated.

Consider the model developed earlier for the multitrait-multimethod situation. For three methods and five traits, the factor pattern and factor correlation pattern have the following form.

$$\Lambda_p = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$
\Phi_P = \begin{array}{c} \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \left[\begin{array}{ccccccccc}
( ) & & & & & & & & \\
0 & ( ) & & & \text{(symmetric)} & & & & \\
0 & 1 & ( ) & & & & & & \\
0 & 1 & 1 & ( ) & & & & & \\
0 & 1 & 1 & 1 & ( ) & & & & \\
0 & 1 & 1 & 1 & 1 & ( ) & & & \\
0 & 1 & 1 & 1 & 1 & 1 & ( ) & & \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & ( ) & \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & ( )
\end{array}\right] \end{array}
$$

If the solution suggests that factors 7 and 9 are closely related, say $\hat{\Phi}_{97} = .98$, then one would collapse these two factors. The collapse is represented by the following patterns in which the factors have been collapsed.

$$
\Lambda_P = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \left[\begin{array}{cccccccc}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{array}\right] \end{array}
$$

$$
\Phi_P = \begin{array}{c}
\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{bmatrix}
(\ ) & & & & & & & \\
0 & (\ ) & & & \text{(symmetric)} & & & \\
0 & 1 & (\ ) & & & & & \\
0 & 1 & 1 & (\ ) & & & & \\
0 & 1 & 1 & 1 & (\ ) & & & \\
0 & 1 & 1 & 1 & 1 & (\ ) & & \\
0 & 1 & 1 & 1 & 1 & 1 & (\ ) & \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & (\ )
\end{bmatrix}
\end{array} .
$$

2. <u>Consistency of solution</u>

The correlations between the factors, (i.e., elements in the Φ matrix) are estimated independently of the other parameters in maximum likelihood factor analysis. Φ must be positive definite (IND≠2) in order that the solution may be viable. That is, the requirements in Φ will be met if all eigenvalues are positive. Computation of such values, however, gives little information about specific variables which may be a basis for a nonpositive condition. Examination of the matrix for elements near $\pm$ 1.0 is helpful insofar as such values reflect troublesome variables. One may also use a heuristic device to delineate inconsistency in relations between variables. The following inequality is such a device.

$$1 - r_{ij}^2 - r_{ik}^2 - r_{jk}^2 + 2(r_{ij})(r_{ik})(r_{jk}) > 0 \qquad i,j,k=1,2, \ldots ,m .$$

The inequality is based on the general condition in regression that the square of the multiple regression coefficient is less than or equal to 1. For examples of the use of the inequality in regression, see Walker and Lev (1953).

Consider an intermediate solution (not illustrated) for Data Set 3, in which the relations among the estimates of factor correlations were not consistent. Substitution of the maximum likelihood estimates for correlations among the three factors yields the following.

$$1 - \hat{\phi}_{12}^2 - \hat{\phi}_{13}^2 - \hat{\phi}_{23}^2 + 2(\hat{\phi}_{12})(\hat{\phi}_{13})(\hat{\phi}_{23})$$

$$= 1 - (.591)^2 - (-.975)^2 - (-.027)^2 + 2(.591)(-.975)(-.027) \not> 0 .$$

Such an inconsistent relation is considered sufficient grounds for rejection or alteration of the model even though other criteria, such as the χ^2, residuals, etc., may suggest that the model is appropriate. The inconsistency occurs rather frequently in solutions, usually in conjunction with other indicators of an inappropriate model, discussed below. In Data Set 3, for example, eight models were hypothesized and eight solutions obtained. Of these eight, five solutions contained $\hat{\Phi}$ matrices some of whose elements were inconsistent. Each case of inconsistency was accompanied by a boundary problem. The boundary problem in each case was confined to the $\hat{\Phi}$ matrix. That is, very large (absolute) values of correlations between factors was obtained.

3. <u>Goodness of fit</u>

Consider the χ^2 statistic discussed earlier.

$$\chi_{\nu}^2 = (N-1) \sum_{q<q'} (A_{qq'} - \hat{\Sigma}_{qq'})^2 / \hat{\psi}_{qq} \hat{\psi}_{q'q'} , \quad q,q' = 1,2 \ldots p$$

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Phi}\hat{\Lambda} + \hat{\psi} = \text{estimated correlation matrix,}$$

$$A = \text{observed correlation matrix.}$$

The function on the right hand side of the equation is an approximation to the function minimized, according to Lawley and Maxwell (1963). The equality requires that the errors be distributed normally and independently and that the sample size be large.

Under null conditions, the data are derived from the model specified, including assumptions and restrictions. The associated χ^2_ν is distributed approximately as given in the appropriate tables. For this research an important type of restriction is the insertion of zero elements as parameters in the factor structure and factor correlation matrices.

The alternative hypothesis is that A is any positive definite matrix of order p x p . That is, no alternative model is specified.

Consider the implications of the statements above for a test of a specific factor analytic model. Suppose one computes a χ^2 value which is significant at some probability level or at least extremely large. The large value may occur for a number of reasons.

For example, the linear model may be inappropriate. That is, it is possible that some nonlinear function of hypothetical factors can describe or explain the data more adequately.

It is also possible that the distributional assumptions are not tenable. If some of the frequency distributions of the observed variables are nonnormal, say skewed badly, then one might obtain a large χ^2 . If basic data are available, one could examine the viability of the normality assumption. However, usually only correlations are presented in published multimethod-multitrait investigations. To the extent that the sample is large, one usually assumes that the author of such publications has ascertained there are no large deviations from normality.

A third possible reason for a significant χ^2 is the incorrectness of the restrictions imposed on the factor structure. In the most simple case, one may require m + 1, rather than m factors to account for the data. This is the aspect of the alternative hypothesis usually stated in texts on factor analysis. It is as proper to reject linearity, or any of the reasons stated here, as it is to reject the hypothesis that there are m common factors. Morrison (1967) provides some discussion on this topic. In addition, the restrictions imposed on the factor structure and factor correlation matrices, may be inappropriate. That is, although some m factor linear model, from the totality of such models, may summarize the data adequately, the restrictions imposed on that model may not be consistent with the data.

One may unambiguously reject a model, but from the χ^2 alone, one cannot surmise the reason for failure of the model. That is, any or all of the items suggested above may account for the model's failure to fit the data. To this extent, the inferences based on the χ^2 are ambiguous.

Consider the right hand side of the χ^2 function, essentially a function of the residuals. The rather important notion that examination of residuals if useful in assessment of models is discussed by Anscombe (1960) and Daniel (1968), for example. However, little systematic research in examination of residuals of factor analytic models appears in the published literature. Rather than interpreting the χ^2 as a terminal decision making device, one may examine the right hand side from the point of view of a descriptive indicator of magnitudes of residuals.

That is, insofar as the optimization procedure effectively minimizes the sum of the ratios indicated, the smaller residuals are associated with smaller specific factors and the larger residuals are associated with larger specific factors. Conditional on specific estimates of ψ , the larger residuals indicate a more inferior fit. Also, the smaller estimates of ψ are associated with larger estimates of the Λ. The lower limit to the index is reached when one has as many factors as measures. That is, the residuals will be exactly equal to zero when the number of (unrestricted) factors hypothesized equals the number of measures. This is equivalent to a principal components analysis. The practical lower limit is higher than zero, however. This is so because one uses the model to obtain a more parsimonious description of the data. This parsimony requires that there be as few factors as possible and fewer factors result in higher residuals. Therefore, one must achieve a compromise between the objective of small residuals and the objective of parsimonious description of the data.

There are some rather important limitations on all of the observations made above. These limitations are a function of the sample size and also of the procedure used to conjecture the models.

To the extent that the sample size is large, the χ^2 approximation is good. That is, one can make significance tests, conditional on a model, with some confidence. No definitive statement regarding just how large the sample must be can be made, however. The sample size in this research is approximately 125 for each of the data sets analyzed.

The large N is also important to the extent that the development of the factor structure from the experimental model requires the assumption.

The procedure used in building the factor analytic models conditional on the preceding models is important. Statements about probability levels of significance tests cannot be made since each test is conditional on the preceding one. Moreover, the expectancy of each χ^2 is not equal to it's degrees of freedom unless one ignores the conditionality of the models.

To this extent, then, the use of significance tests in the usual sense is fatuous. There seems to be no general approach to this sequential hypothesis testing problem in factor analysis currently available.

The following heuristic approach constitutes a partial resolution of the dilemma. There seems to be no good reason to discount use of the function computed as one index of the extent to which the model fails to fit the data. Other indices, to be considered simultaneously, are the boundary conditions and consistency of solution described above. The index may be evaluated with respect to its algebraic lower limit of zero. It may be evaluated with respect to an upper limit of, say, the χ^2 and expectancy for the very first model hypothesized. Statements about rejection on the basis of a significance tests and the expectancies are viable for the first model, since it is not conditional on any preceding models. The magnitude of such a value gives one some idea of what type of fit is possible and provides an unambiguous χ^2 value against which other computed values can be compared. The relative magnitudes of the computed values can be manipulated in the context of the reasons suggested above for failure of the models.

4. Other considerations

Several other characteristics of the solution are informative. The RMLFA Computer Program provides an interval estimate of the parameters. The first order partial derivatives at the estimated maximum of the likelihood function are also obtained.

The former is an approximation and provides one with information on how well the maximum for each parameter is determined. If the function is flat, then the points of inflection will be widely spaced and the (approximate) confidence interval will be large. If a well defined minimum (maximum) is located, then the points of inflection will be closely spaced.

The first derivatives with respect to all the parameters provide one with information on the flatness of the function at the point for which the estimates of parameters are provided. If the solution is a viable one, the first derivatives should be small. The RMLFA Computer Program automatically terminates the iterative procedure when the derivatives approach .0005, and the user is informed of this through the program output message IND = 0. The value .0005 is rather restrictive and may be unreasonable in terms of accuracy of the computer and available computing time. Therefore, a higher value at which to terminate the iterations may be chosen. The convenient specification is to limit the number of iterations, the number specified being based on previous analyses and the size of the derivatives obtained therein. The program submits a message informing the operator that the specified maximum number of iterations has been reached (IND = 4).

It can occur that the data are such that no maximum likelihood solution is possible, conditional on some factor analytic model. Moveover, one can investigate the possibility of solution in the context of explicit criteria to which the correlation matrix (A) must conform. In order to obtain $\underline{a}$ solution with, say, m factors, the data must conform to the following theorem, taken from Anderson and Rubin (1956). A necessary and sufficient condition that Σ be a covariance matrix of a factor analysis model with m factors, is that there exists a diagonal matrix, ψ, with nonnegative elements such that $\Sigma - \psi$ is positive semidefinite, and of rank m. This means that Σ, Φ and ψ must be positive definite in order for the solution to fall within the allowable region of the vector space defined by the model. With actual data, one may obtain an A which is not positive definite because of an inappropriate model or an unusual event regarding the sample or an error in computations.

In fact, the question of existence of the solution is less important than it may seem. One either obtains or does not obtain a solution, conditional on a model. If one does not obtain a solution, then one changes the model in some respect. Changes which may be made in the model are fully discussed below..

A relevant criteria not considered in detail is the computed χ^2 divided by its degrees of freedom. The quantity is an estimate of the residual mean square adjusted for degrees of freedom. It is, in effect, an unbiased index of the magnitude of residuals and may well be more appropriate than the simple χ^2 statistic. Unfortunately, not enough time was available to fully explore this index.

IV. DATA ANALYSIS

A description and tabulated results of analyses of three sets of data are presented. An interpretation of the results with respect to a functional frame of reference is also provided.

Data Sets 1, 2, and 3 consist of information obtained from Sociology, Clinical Psychology and Industrial Psychology, respectively. The data are given in tables at the end of each section describing the interpretation of the analysis. A summarization of the interpretation appears in the latter portion of each section.

Table 14 contains a summarization of the model building process required in the analysis of each data set. The table provides documentation for the comments made in the earlier discussion of rationale for changes in the models.

A. Data Set 1

1. Description

Data Set 1 is based on a study by Borgatta (1955). The purpose of the research was to assess group interaction processes under specified conditions. The subjects consisted of 126 Air Force personnel, who were given precisely defined tasks to perform in three man groups. Observations were obtained under three modes (methods of behavior: Actual (Free) Behavior, Role Playing, and Projective Tests. Ratings from independent observers were obtained and pooled to obtain scores on the following traits.

1. Shows Solidarity - raises others' status, gives help, reward.
2. Gives Suggestion - direction, implying autonomy for others.
3. Gives Opinion - evaluation analysis, expresses feeling, wishes.
4. Gives Orientation - information, repetition, confirmation.
5. Shows Disagreement - shows passive rejection, formality, withholds help.

More precise descriptions of tasks, traits and methods are given in the reference cited. The correlations appear in Table 3, and the Λ and Φ matrices appear in Tables 4 and 5 respectively. Table 6 contains the variance components estimates.

2. Interpretation

Consider the first column of the $\hat{\Lambda}$ matrix which appears to be a general factor. The most substantial contribution to this factor is the trait "Shows Solidarity", with loadings of .58, .88, .24 for the respective methods. "Showing Disagreement" also makes up this factor, with loadings of .34, .62, and .24. The other trait-method combinations contribute somewhat less than these two. These observations suggest that Showing Solidarity is not really a distinct trait, but is a part of a more global phenomena. Showing Disagreement is strongly related to showing Solidarity according to the correlation between the factors corresponding to these traits. That is, $\hat{\Phi}_{95} = .52$. Therefore, comments which apply to Showing Solidarity also apply to Showing Disagreement, to the extent that their factors are linearly related. These observations are further supported by those below.

Column 2 is a method factor, "Free Behavior". Strong bias due to this method is apparent in measurement of the traits Giving Suggestion, Giving

Opinions and Showing Disagreement. One reaches this conclusion by noting the loadings of .56, .59 and .45 for the respective traits in column 2. Using this method, apparently the best trait to measure is "Gives Orientation" because method bias is small (loading of -.08) relative to the others. Moreover, the trait-variable combination is loaded heavily on it's own factor. This is, in column 8, the trait loads highest (.94) for this method, the other lower loadings in that column indicating poorer measurement.

The second method factor, represented in column 3, reflects biases due to measurement by means of "Role Playing". The loading of .87 suggests that Giving Opinions is badly biased by using this method. However, it appears that one can measure Giving Suggestions well, without being mislead by such bias. The same is true for Giving Orientation but the evidence is not quite as strong. The loadings of -.01 and .10 in the second column, and the corresponding loadings of .89 and .74 in the 6th and 8th columns support the preceding observations. The latter set of loadings also suggest the discriminability of the traits. The discriminability is undermined a bit by the correlation ($\hat{\varrho}_{96}$ = -.23) between the factors but this does not appear to be serious.

Column 4, the last method factor, represents Projective Testing and loadings on this factor of .41, .79, .62, .54 and .55 are rather high. Moreover, the general factor loadings (first column) corresponding to this method are low, .24, .09, .36, .05 and .24. This suggests that the projective tests do not measure a "global factor" so much as they measure the method of giving projective tests. Much of the variance is attributable to method rather than trait. Discrimination of traits, using this method is impossible if one considers the last five rows of columns 5,6,7,8 and 9. The Projective Test method cannot reflect distinctive traits in any consistent way, at least for this group of traits.

One can summarize the statements made above. The general factor, comprised mainly of the trait called Showing Solidarity, pervades all measures to some extent. Giving Orientation is best measured using the method of Free Behavior, insofar as method bias is rather small in this case. Giving Suggestion is best measured by Role Playing for the same reason. Gives Opinion is subject to method bias in any situation considered here, but such bias is least when the method of Free Behavior is used. The worst method to use, with respect to measurement of these distinct traits, is Projective Tests. The traits with the least discriminant validity are Shows Solidarity and Shows Disagreement. They do not appear to be discriminable from a general, global type of assessment.

The variance components estimates are, in a sense, a condensation of the $\hat{\Lambda}$ and $\hat{\Phi}$ matrices and of the interpretive comments made above. Table 6 contains the relevant data. The interpretation of such components supports the interpretive statements about $\hat{\Lambda}$ and $\hat{\Phi}$, although in less detail. That is, the variance component for measure 4I suggests that the trait of Gives Orientation as measured by Free Behavior is functionally viable. A large part of the total variance is attributable to the trait factor. Analogous interpretations can be devised for the components associated with variables 2II and 3I. The projective tests (method III) fare poorly and appear to be inconsistent with a simple model such as this. Error variance is rather high for this method.

Table 3. Correlations (A) among ratings of five traits in three modes of behavior (N = 125), Data Set 1

j	k	1 I	2 I	3 I	4 I	5 I	1 II	2 II
1	I	100						
2	I	25	100					
3	I	13	24	100				
4	I	-14	26	52	100			
5	I	34	41	27	02	100		
1	II	43	43	08	10	29	100	
2	II	16	32	00	24	07	37	100
3	II	15	27	06	38	12	01	10
4	II	-12	24	44	74	08	04	18
5	II	51	36	14	-12	50	39	27
1	III	20	17	16	12	08	17	12
2	III	05	21	05	08	13	10	19
3	III	31	30	13	-02	26	25	19
4	III	-01	09	30	35	-05	03	00
5	III	13	18	10	14	19	22	28

[j] Trait: 1, Shows solidarity; 2, Gives suggestion; 3, Gives opinion ; 4, Gives Orientation; 5, Shows Disagreement.

[k] Method: I, Free Behavior; II, Role Playing; III, Projective Tests.

3 II	4 II	5 II	1 III	2 III	3 III	4 III	5 III
100							
40	100						
23	-11	100					
30	17	22	100				
-20	06	30	32	100			
15	-04	53	31	63	100		
19	33	00	37	29	32	100	
02	04	23	27	51	47	30	100

(symmetric)

Table 4. Factor loadings ($\hat{\Lambda}$) for Data Set 1

j	k	A	C_I	C_{II}	C_{III}	B_1	B_2	B_3	B_4	B_5
1	I	58	25	00	00	29	00	00	00	00
2	I	40	56	00	00	00	27	00	00	00
3	I	04	59	00	00	00	00	57	00	00
4	I	-09	-08	00	00	00	00	00	94	00
5	I	34	45	00	00	00	00	00	00	31
1	II	88	00	-09	00	-35	00	00	00	00
2	II	40	00	-01	00	00	89	00	00	00
3	II	09	00	87	00	00	00	-23	00	00
4	II	-11	00	10	00	00	00	00	74	00
5	II	62	00	31	00	00	00	00	00	71
1	III	24	00	00	41	49	00	-06	47	-16
2	III	09	00	00	79	-20	31	09	-08	37
3	III	36	00	00	62	24	11	-01	23	28
4	III	05	00	00	54	43	-07	28	43	-30
5	III	24	00	00	55	-02	28	12	-00	09

Table 5. Correlations between factors ($\hat{\Phi}$) for Data Set 1

Chi-square = 18.09, df = 12

A	100								
1	00	100							
2	00	33	100						
3	00	18	03	100		(symmetric)			
4	00	-37	02	-11	100				
5	00	-17	13	-17	-08	100			
I	00	-23	-04	-18	13	15	100		
II	00	48	64	03	-70	33	52	100	
III	00	09	-15	06	51	-23	04	-36	100

Table 6. Variance components for Data Set 1

j	k	$\hat{A}^2_{jk}$	$\hat{B}^2_{1k}$	$\hat{B}^2_{2k}$	$\hat{B}^2_{3k}$	$\hat{B}^2_{4k}$	$\hat{B}^2_{5k}$	$\hat{C}^2_{jk}$	$2\Sigma\hat{B}_{jk}\hat{C}_{jk}\hat{\phi}_{jk}$	$\hat{\sigma}^2_{\epsilon_{jk}}$	$\hat{Var}(y_{i(jk)})$
1	I	34	08					06	-05	57	100
2	I	16		07				30	-05	50	99
3	I	00			33			35	-16	48	100
4	I	01				88		01	-07	17	100
5	I	12					10	20	-02	55	99
1	II	76	12					01	00	09	98
2	II	16		79				00	00	05	100
3	II	01			05			76	-01	15	98
4	II	01				55		01	09	33	99
5	II	38					50	10	-06	07	99
1	III	05	24	00	00	22	03	17	-41	70	100
2	III	01	04	10	01	01	14	61	-20	21	93
3	III	12	06	01	00	05	08	37	-06	36	97
4	III	00	19	01	08	19	09	29	-29	41	97
5	III	05	00	08	01	00	01	29	-06	60	98

B. Data Set 2

1. Description

Clinical psychology students ($N = 124$) rated themselves and three teammates on each of five traits or characteristics. The median of the three teammate ratings was used as the score for this method of measurement. An assessment staff also rated the Ss , the ratings of the three members of the staff being pooled for a score on a given S . Three methods of measurement are then, Staff ratings, Teammate ratings and Self ratings. The Dimensions or traits rated are: Assertive, Cheerful, Serious, Unshakeable Poise, and Broad Interests. These traits were deliberately chosen from a 66 x 66 matrix of traits measured using the methods described. The criterion for considering a trait was essentially its validity and the selection was made by Campbell and Fiske (1959). Full descriptions of the traits, methods and data are contained in Kelly and Fiske (1951). The model assumes that each psychologist has 9 true scores: five trait scores, three scores reflecting method bias and one over-all score reflecting the ratees over-all reputation. Method bias occurs because each rater perceives the ratees in a different situation, thus the over-all evaluation of the ratee may depend on the source of the rating.

The correlations among the 15 measures are reported in Table 7 and two solutions are reported in Tables 8 and 9 and Tables 10 and 11.

2. A Poor Solution

It may be informative to consider a solution which is poor relative to the criteria established earlier for rejection of a model. Such a solution is given in Table 8 and 9 for Data Set 3. The last three variables in the $\hat{\Lambda}$ matrix are correlated in such a way as to produce a negative definite $\hat{\Phi}$ matrix. That is, the following inequality should hold if one were to accept the solution.

$$1 - \Phi^2_{78} - \Phi^2_{19} - \Phi^2_{89} + 2(\Phi_{78})(\Phi_{79})(\Phi_{89}) > 0 \quad .$$

Substitution of the correlations between the variables yields

$$1 - (.54)^2 - (-.98)^2 - (-.02)^2 + 2(.54)(-.98)(-.02) < 0 \quad .$$

In addition, earlier solutions suggested that one could constrain the methods to be independent of traits. In other words, the entries in the $\hat{\Phi}$ matrix, corresponding to the correlations between these two types of variables would be equal to zero. This is illustrated in the table The solution also suggests that factors 7 and 9 (representing Methods I and III) are not different. That is, the absolute magnitude of the correlation between them ($\hat{\Phi}_{79} = -.98$) is high enough to warrant altering the model so that only one, factor, representing both methods, is hypothesized. This alteration is implemented by collapsing one of these factors into the other. Specifically, the nonzero loadings which were hypothesized to be in specific locations in factor 9 are hypothesized to be in the corresponding positions in factor 7. Factor 9 is thus eliminated.

Intermediate solutions (not illustrated) suggested that Traits 1 and 2 are not different. These two factors were then collapsed into one.

Imposing the constraints suggested by the preceding observations, solution 2 is reported in Tables 10 and 11. This solution is satisfactory in that the minima occur in the acceptable region and the chi-square is small. However, the solution for the second derivatives could not be determined because $\hat{\Sigma}$ is not positive definite thus we can not, for any given estimate, evaluate how well it is determined.

3. Interpretation

Rather than examining the results with respect to convergent and discriminant validity, one might choose to take a more functional orientation. That is, an emphasis is placed on the determination of best ways to measure given traits and the relations between them, as well as assessing biases attached to different methods.

The interpretation of the results of Data Set 3 might be best organzed if one considers the $\hat{\Lambda}$ matrix a column at a time.

The first column appears to represent a general reputation factor, into which the variables enter with differing importance. From the loadings of .78, .49 and .48 in the first column of Λ, it appears that all three types of raters agree that Cheerfulness is associated with general reputation. However, the staff and ratees feel that Unshakeable Poise contributes to general reputation (loading of .41 and .28) while the peers feel it is negatively associated with Seriousness (-.32).

The second column is defined primarily by Trait 1, Assertiveness, and secondarily by Trait 2, Cheerfulness. Note the loadings of .83, .85 and .53 for Assertiveness and the loadings of .36, .43 and .11 for Cheerfulness. This suggests that Cheerfulness, per se, is not a distinct trait. Instead, all three types of raters confuse this with Assertiveness. From the $\hat{\Phi}$ matrix, one can see that Assertiveness correlates negatively with Seriousness ($\hat{\Phi}_{32} = -.33$), and positively with the last two traits, Unshakeable

Poise ($\hat{\Phi}_{42} = .39$) and Broad Interests ($\hat{\Phi}_{52} = .57$) .

Trait 3, Seriousness, is measured independently of the other traits. That is, its correlation with the other factors is almost zero, largest correlations being $\hat{\Phi}_{43} = .11$, $\hat{\Phi}_{53} = .02$. The best measure of this trait comes from Staff ratings, since the staff does not regard seriousness as part of overall reputation or, for that matter, part of any other factor. That is, the most pure (therefore the least biased) measure of seriousness is obtained from Staff Ratings.

Trait 4, Unshakeable Poise, defines the fourth column and is well measured by the Staff Ratings and relatively poorly measured by the Self Ratings. This is apparent from the loading of .71 (Staff) versus .29 (Teammate and Self) on the fourth factor. The loading of .64 in the sixth column, a method factor, supports the preceding observation insofar as it suggests a high method bias in self evaluation of Unshakeable Poise.

Trait 5, Broad Interests, defines the fifth column and also seems to be best measured through Staff Ratings. The loading of .71 on that factor and the loading of -.26 on the corresponding method factor (column 6) suggest that Broad Interests are being measured with little bias relative to the other methods. The inferiority of the Teammate Ratings is shown by the .65 loading in the fifth column and .35 in the corresponding method bias factor (column 7). This method is still superior to the Self Ratings, however, this last observation is shown by the loading of .62 on the trait factor and of .64 on the method factor.

One can summarize the observations above in the following manner. The trait of Cheerfulness is not distinct as measured by the methods used. In fact, it is associated with general reputation and assertiveness. The trait of Assertiveness appears to be distinct and discriminable from other traits. It is best measured by Teammate Ratings with Staff Ratings being a very close second. Seriousness is best measured by the Staff Ratings. Unshakeable Poise is also best measured by Staff Ratings and the worst method to use is the Self Ratings, which have a large method bias. The trait of Broad Interests is best measured by Staff Ratings, although Teammate Ratings are only a little inferior.

A comparison of these results with those conclusions reached by Campbell and Fiske (1959) seems appropriate. Briefly, they suggest that Assertive has very good validity and Cheerful, Serious and Broad Interests have validities based on less convincing evidence. Unshakeable Poise is said to have almost trivial evidence for validity. The conclusions presented earlier, based on the factor analytic solution, are consistent with the preceding observations but contain additional information and are much more specific, of course. Moreover, by looking at the highest and lowest loadings for 3 types of ratings of the same trait in the methods factor columns (6 and 7), it is apparent that Unshakeable Poise contains the most biases and Assertive fares best in this respect.

The variance components estimates effectively summarize these observations. That is, from the magnitudes of the components relative to one another, one can make inferences about the factors to which the components are related. Specifically, the trait, Assertive, is best measured by Staff Ratings and Teammate Ratings, insofar as these methods produce consistent, unbiased assessments. Cheerful is a global trait, rather than being distinct from the other traits. Serious and Unshakeable Poise are best measured by Staff Ratings.

Table 7. Correlations (A) among ratings from assessment study of clinical psychologists (N = 124), Data Set 2

j k	1 I	2 I	3 I	4 I	5 I	1 II	2 II
1 II	100*						
2 I	37	100					
3 I	-24	-14	100				
4 I	25	46	08	100			
5 I	35	19	09	31	100		
1 II	71	35	-18	26	41	100	
2 II	39	53	-15	38	29	37	100
3 II	-27	-31	43	-06	03	-15	-19
4 II	03	-05	03	20	07	11	23
5 II	19	05	04	29	47	33	22
1 III	48	31	-22	19	12	46	36
2 III	17	42	-10	10	-03	09	24
3 III	-04	-13	22	-13	-05	-04	-11
4 III	13	27	-03	22	-04	10	15
5 III	37	15	-22	09	26	27	12

[j]Trait: 1, Assertive; 2, Cheerful; 3, Serious; 4, Unshakable poise; 5, Broad Interests.

[k]Method: I, Staff ratings; II, Teammate ratings; III, Self ratings.

*Decimal points omitted.

3 II	4 II	5 II	1 III	2 III	3 III	4 III	5 III
			(symmetric)				
100							
19	100						
19	29	100					
-15	12	23	100				
-25	-11	-03	23	100			
31	06	06	-05	-12	100		
00	14	-03	16	26	11	100	
-07	05	35	21	15	17	31	100

Table 8. Factor loadings ($\hat{\Lambda}$) for Data Set 2 (Solution 1)

j	k	A	B_1	B_2	B_3	B_4	B_5	C_I	C_{II}	C_{III}
1	I	68	56	00	00	00	00	-15	00	00
2	I	72	00	50	00	00	00	-02	00	00
3	I	-27	00	00	52	00	00	29	00	00
4	I	50	00	00	00	38	00	31	00	00
5	I	44	00	00	00	00	53	31	00	00
1	II	62	54	00	00	00	00	00	11	00
2	II	63	00	19	00	00	00	00	26	00
3	II	-43	00	00	53	00	00	00	38	00
4	II	03	00	00	00	18	00	00	54	00
5	II	28	00	00	00	00	51	00	54	00
1	III	49	24	00	00	00	00	00	00	20
2	III	33	00	36	00	00	00	00	00	23
3	III	-17	00	00	46	00	00	00	00	25
4	III	16	00	00	00	73	00	00	00	42
5	III	34	00	00	00	00	53	00	00	53

Table 9. Correlations between factors (Φ) for Data Set 4 (Solution 1)

Chi-square = 47.28, df = 47

A	100								
1	00	100							
2	00	-40	100				(symmetric)		
3	00	08	04	100					
4	00	-08	42	30	100				
5	00	25	-42	36	05	100			
I	00	00	00	00	00	00	100		
II	00	00	00	00	00	00	54	100	
III	00	00	00	00	00	00	-98	-02	100

Table 10. Factor loadings (Λ) for Data Set 2 (Solution 2)

j	k	A	B_1 & B_2	B_3	B_4	B_5	C_I & C_{III}	C_{II}
1	I	10	83	00	00	00	15	00
2	I	78	36	00	00	00	03	00
3	I	-07	00	65	00	00	-28	00
4	I	41	00	00	71	00	-23	00
5	I	04	00	00	00	71	-26	00
1	II	04	85	00	00	00	00	07
2	II	49	43	00	00	00	00	34
3	II	-32	00	62	00	00	00	26
4	II	-08	00	00	29	00	00	57
5	II	-08	00	00	00	65	00	35
1	III	17	53	00	00	00	14	00
2	III	48	11	00	00	00	23	00
3	III	-18	00	41	00	00	26	00
4	III	28	00	00	29	00	40	00
5	III	01	01	00	00	62	64	00

Table 11. Corelations between factors (Φ) for Data Set 2 (Solution 2)

Chi-square = 46.67, df = 53

A	100						
1,2	-00	100			(symmetric)		
3	00	-33	100				
4	00	39	11	100			
5	00	57	02	50	100		
I,III	00	00	00	00	00	100	
II	00	00	00	00	00	-13	100

Table 12. Variance components for Data Set 2

j	k	$\hat{A}^2_{jk}$	$\hat{B}^2_{jk}$	$\hat{C}^2_{jk}$	$2\hat{B}_{jk}\hat{C}_{jk}\hat{\Phi}_{jk}$	$\hat{\sigma}^2_{\epsilon_{jk}}$	$Var(\hat{y}_{i(jk)})$
1	I	.01	.69	.02	00	.27	.99
2	I	.61	.13	00	00	26	1.00
3	I	.01	.43	.08	-01	48	.99
4	I	.16	.50	.05	01	25	.97
5	I	.00	.50	.07	00	42	.99
1	II	.00	.72	.01	00	27	1.00
2	II	.24	.19	.12	-01	46	1.00
3	II	.10	.38	.07	00	45	1.00
4	II	.01	.08	.33	00	57	1.00
5	II	.01	.42	.12	00	44	.99
1	III	.03	.28	.02	00	68	1.01
2	III	.23	.01	.05	-01	71	.99
3	III	.03	.17	.07	00	74	1.01
4	III	.08	.08	.16	00	68	1.00
5	III	.00	.38	.41	01	24	1.04

C. Data Set 3

1. Description

Data Set 3 consists of information obtained by MacKinney (1968) in an industrial study of managerial performance. Department heads (N = 124) at industrial installations were rated by their superiors (plant managers) and also by two subordinates (foremen), on six attributes. The two foremen differed with respect to the ratings given to the foremen themselves by the department head in a previous study. That is, the department head had earlier assessed one foreman as being high on certain attributes and the other as being low on the same attributes. The methods of rating department heads, then, are (1) plant manager, (2) foreman with high ratings and (3) foreman with a low rating. The three methods correspond to three different viewpoints of the department head and one might expect some biases in judgement attributable to the different viewpoints. The six attributes or traits of the department head, judged by each of the methods of judgement were:

1) Intellectual traits
2) Human relations
3) Concern for quality
4) Leadership orientation
5) Independence
6) Achievement orientation.

The data and solutions appear in Table 13 and Tables 14 and 15, respectively.

Further details concerning the nature of the methods and traits are available in MacKinney (1968).

2. Interpretation of the solution

Consider the first column of the $\hat{\Lambda}$ matrix, representing the first factor common to all observations. From the factor correlations, the $\hat{\Phi}$ matrix, one may note that this factor is independent of the remaining factors. The latter, of course, are alleged to explain the communality not attributable to the first factor. The rather high loadings in the first column define the nature this factor: $\hat{\Lambda}_{21} = .36$, $\hat{\Lambda}_{81} = .78$, $\hat{\Lambda}_{14,1} = .47$.

Specifically, it appears that the factor is a general rating which reflects mostly human relations ability. The factor is least associated with achievement orientation. The presence of this general factor suggests that all observations have in common a global assessment, rather than assessment of specific traits. It appears that these comments are true for each type of rater. That is, plant managers, high-rated foremen and low-rated foremen each weight human relations ability most and achievement orientation least in evaluating any aspect of the subject. This is apparent from the high and low loadings within each set of six measures.

The next three columns represent factors which are common to the measures and which are attributable to methods biases. That is, column 2 represents a factor which relflects the extent of biases in plant manager judgements. Columns 3 and 4 represents analogous methods factors for the two types of foremen. From the $\hat{\Phi}$ matrix, one can see that these three factors are nearly independent. In fact, the low homogeneous correlations ($\hat{\phi}_{32} = .32$, $\hat{\phi}_{42} = .36$, $\hat{\phi}_{43} = .22$) suggest that there is a small factor common to each type of bias. For the most part however, the biases are more nearly independent than they are related. Note that all the loadings on each of these three factors are extremely high, the range of $\hat{\Lambda}$'s being .50 to .90. This means that the judgement is more a function of the raters making a judgement than it is a function of the trait being judged. Moreover, the trait which is the object of attention is unimportant insofar as judgements are made about general human relations rather than the specific traits. Considering the magnitude of loadings on methods in comparison to the general faction loadings, the least biased method of assessing the general facfors is to obtain judgements of human relations abilities from the high rated foreman.

The results may be summarized as follows. One cannot obtain unequivocal judgements about these six specific traits of department heads using the methods indicated. One can only obtain a global judgement which is mainly a reflection of the department heads' ability to deal with people generally. Any judgement, regardless of whether it is made by superiors or subordinates, is heavily biased and the bias is added to the observation. The magnitudes of the bias appears to be much the same for the three types of raters. The superiors and subordinates agree to the extent that the general factor explains some commonness among observations. The best way to assess human relations ability appears to be administration of the corresponding questionnaire to foreman rated high in previous assessments.

Examination of the variance components suggests, more concisely, that the biases in judgements by each of the three types of raters are extensive. Moreover, there is considerable error attached to the measures. The general factor, associated with Human Relations, appears to be least biased when judgements are obtained from foremen previously rated high by their superiors.

Table 13. Correlations (A) among ratings of six traits using three methods of assessment (N = 124), Data Set 3

j	k	1I	2I	3I	4I	5I	6I	1II	2II	3II	4II	5II	6II	1III	2III	3III	4III	5III	6III
1	I	100																	
2	I	38	100																
3	I	54	42	100															
4	I	65	53	61	100							(symmetric)							
5	I	70	36	62	73	100													
6	I	63	30	63	68	83	100												
7	II	24	16	17	22	21	08	100											
2	II	19	32	13	25	23	08	74	100										
3	II	20	18	31	31	26	18	75	61	100									
4	II	30	27	24	31	32	16	83	86	73	100								
5	II	25	25	27	24	26	11	75	50	73	77	100							
6	II	27	21	32	32	30	21	75	62	74	77	80	100						
1	III	29	21	23	35	34	24	22	25	20	22	26	26	100					
2	III	26	33	13	30	26	16	33	44	22	38	33	30	72	100				
3	III	29	19	20	37	34	28	19	20	18	21	21	23	83	68	100			
4	III	35	23	23	37	37	30	28	31	21	31	31	29	84	81	82	100		
5	III	21	10	16	30	29	21	211	16	15	17	23	21	79	67	79	79	100	
6	III	21	11	11	31	28	23	12	10	07	11	20	16	78	61	82	74	86	100

Table 14. Factor loadings ($\hat{\Lambda}$) for Data Set 3

j	k	A	C_I	C_{II}	C_{III}
1	I	15	74	00	00
2	I	36	42	00	00
3	I	00	72	00	00
4	I	15	81	00	00
5	I	12	91	00	00
6	I	-04	89	00	00
1	II	47	00	75	00
2	II	78	00	50	00
3	II	27	00	80	00
4	II	64	00	70	00
5	II	40	00	77	00
6	II	28	00	84	00
1	III	17	00	00	89
2	III	47	00	00	72
3	III	11	00	00	90
4	III	29	00	00	88
5	III	05	00	00	90
6	III	-01	00	00	90

Table 15. Correlations between factors ($\hat{\Phi}$) for Data Set 3

Chi-square = 168.61, df = 114

A	100		(symmetric)	
I	0	100		
II	0	32	100	
III	0	37	22	100

Table 16. Variance components for Data Set 3

j	k	$\hat{A}_{jk}$	$\hat{C}_{jk}$	$\hat{\sigma}^2_{\epsilon_{jk}}$	$\hat{Var}(y_{i(jk)})$
1	I	02	55	42	99
2	I	13	18	70	101
3	I	00	52	50	102
4	I	02	66	33	101
5	I	01	83	16	100
6	I	01	79	21	102
1	II	22	56	22	100
2	II	61	25	14	100
3	II	07	64	30	101
4	II	41	49	10	100
5	II	16	59	26	101
6	II	08	71	22	101
1	III	03	79	18	100
2	III	22	52	25	99
3	III	01	81	18	100
4	III	08	77	15	100
5	III	00	81	20	101
6	III	00	81	21	102

Table 17. Summary information on solutions obtained in analysis of Data Sets 1,2, and 3

Data Set	Model	IND	Bound Problem	Consis-tency	d.f.	Computed χ^2
1	A	2	$\bar{\Phi}$	No	62	143.44
	B	2	$\bar{\Phi}$	No	65	136.85
	C	2	$\bar{\Phi}$	No	68	145.24
	D	4	$\bar{\Phi}$	No	71	146.23
	E	0	None	Yes	73	143.25
	F	2	$\bar{\Phi}$	No	47	100.62
	G	2	$\bar{\Phi}$	No	54	126.74
	H	2	$\bar{\Phi}$	No	60	150.82
	I	2	$\bar{\Phi}$	No	47	136.23
	J	4	None	Yes	66	148.56
	K	2	ψ	Yes	66	147.42
	L	2	ψ	Yes	64	144.95
	M	2	ψ	Yes	40	55.84
	N	0	None	Yes	81	192.56
	O	2	$\bar{\Phi}$	No		
	P	2	ψ	Yes	51	105.81
	Q	2	ψ, Λ	Yes	32	58.02
	R[1]	4	None	Yes	32	1.27
	S	2	$\bar{\Phi}$	Yes	42	.42
	T	2	$\bar{\Phi}$	No	42	105.30
	U	2	$\bar{\Phi}$	No	30	95.07
	V	2	$\bar{\Phi}$	No	28	85.22
	W	4	$\bar{\Phi}$	Yes	35	54.74
	X	2	$\bar{\Phi}$	No	41	105.75
	Y	2	$\bar{\Phi}$	No	41	65.60
	ZA	2	4	Yes	12	17.40
	ZB	4	None	Yes	12	18.09

Table 17. (continued)

Data Set	Model	IND	Bound Problem	Consis-tency	d.f.	Computed χ^2
2	A	2	Φ	No	62	59.81
	B	2	Φ	No	64	111.67
	C	2	Φ	No	47	49.15
	D	2	Φ	No	47	47.28
	E	4	Λ	Yes		
	F	2	None	Yes	56	52.51
	G	2	Φ	No	49	48.23
	H	0	None	Yes	53	46.67
3	A	2	Φ	No	63	18.43
	B	2	None	No		
	C	4	None	No	105	155.43
	D[2]	2	None	Yes	114	168.61

[1]This solution, though satisfactory, resulted in all loadings being near zero on some factors. This suggested that these factors could be eliminated.

[2]Although a possible error condition (IND=2) is indicated, it is not crucial enough to warrant rejection of model.

V. ALTERNATIVE STRATEGIES

There are a number of viable approaches to estimation of parameters associated with a model for variance-covariance matrices. The development of one such approach, meeting with reasonable success, has been described above. Two other possible methods are described in this section. Both are discussed with respect to the multimethod-multitrait data matrix.

The first approach is based on a paper by Wolins (1964). It is an amplification and expansion of the information which that paper contains. Some exploratory research has been completed, since this approach is actually the first one tried. It was set aside, however, for the sake of expediency. That is, although the approach is a viable one, it presents some problems with respect to implementation. Computer programming and derivations of some equation systems were required. Such tasks were completed but a more easily implemented analysis was possible when the RMLFA Computer Program became available. Further clarification is provided below.

The second approach described is a summary of a paper be Bock and Bargmann (1966), which appears to be relevant in several senses. The rationale and methodology is similar to that which led to a solution in this dissertation. This approach was not explored because of the difficulty of its implementation, and the limited time available to the writer.

A. Method 1

One can conjecture that the model which summarizes the multitrait-multimethod data is of the form

$$Y_{i(jk)} = B_{jk}X_{ij} + C_{jk}X_{ik} + \epsilon_{i(jk)}$$

$$X_{ij} \ , \ X_{ik} \sim N(0,1)$$

$$\epsilon_{i(jk)} \sim NID(0, \sigma^2_{e_{jk}}) \ .$$

This model, as stated, is identical in form and definition of terms, to Equation 3.11. However, the general factor represented by the term $A_{jk}x_i$ in Equation 3.11, is absent in this case.

In a manner similar to the earlier development, let

$\rho^{BB}_{jj'}$ = correlation between the hypothetical (trait) variables X_{ij} and $X_{ij'}$, $j, j' = 1, 2, \ldots, b$,

$\rho^{CC}_{kk'}$ = correlation between the hypothetical (method) variables X_{ik} and $X_{ik'}$, $k, k' = 1, 2, \ldots, c$.

Suppose one assumes that the traits do not correlate with the methods. In addition, one may choose to restrict the correlations $\rho^{BB}_{jj'}$, $\rho^{CC}_{kk'}$ in the following manner:

$$\rho^{BB}_{jj'} = D_j D_{j'} \quad , \quad \rho^{CC}_{kk'} = E_k E_{k'} \quad .$$

That is, the correlation between traits or methods may be explained as a multiplicative function, depending on the correlated variables.

The elements of the variance-covariance matrix can be represented as

$$E\{Cov(y_{i(jk)}, y_{i(j'k')}) = B_{jk}B_{j'k'}D_j D_{j'} + C_{jk}C_{j'k'}E_k E_{k'} + \sigma^2_{e_{jk}} \delta_{jj'} \delta_{kk'} \quad .$$

where

$$D_jD_{j'} = \rho^{BB}_{jj'}$$

$$E_kE_{k'} = \rho^{CC}_{kk'}$$

$$\delta_{jj'}, \delta_{kk'} = \text{Kronecker deltas.}$$

One can show that the matrix whose elements are represented by the equation above can be described in terms of a factor analytic model. That is, for the case of three methods (k=1,2,3,) and three traits (j=1,2,3,) we have the following. The Λ matrix of unknowns is given in Table 18.

$$\psi = \begin{bmatrix} \sigma^2_{e_{1I}} & & & & & & & & \\ 0 & \sigma^2_{e_{2I}} & & & & \text{(symmetric)} & & & \\ 0 & 0 & \sigma^2_{e_{3I}} & & & & & & \\ 0 & 0 & 0 & \sigma^2_{e_{1II}} & & & & & \\ 0 & 0 & 0 & 0 & \sigma^2_{e_{2II}} & & & & \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{e_{3II}} & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{e_{1III}} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{e_{2III}} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{e_{3III}} \end{bmatrix}$$

Σ = expectation of variance-covariance matrix $\Sigma = \Lambda\Lambda' + \psi$.

Table 18. Λ matrix

$$
\Lambda = \begin{bmatrix}
B_{1I} & D_{I} & C_{1I} & E_{1} & B_{1I}\sqrt{1-D_{I}^{2}} & 0 & 0 & C_{1I}\sqrt{1-E_{1}^{2}} & 0 & 0 \\
B_{2I} & D_{I} & C_{2I} & E_{2} & B_{2I}\sqrt{1-D_{I}^{2}} & 0 & 0 & 0 & C_{2I}\sqrt{1-E_{2}^{2}} & 0 \\
B_{3I} & D_{I} & C_{3I} & E_{3} & B_{3I}\sqrt{1-D_{I}^{2}} & 0 & 0 & 0 & 0 & C_{3I}\sqrt{1-E_{1}^{2}} \\
B_{1II} & D_{II} & C_{1II} & E_{1} & 0 & B_{1II}\sqrt{1-D_{II}^{2}} & 0 & C_{1II}\sqrt{1-E_{1}^{2}} & 0 & 0 \\
B_{2II} & D_{II} & C_{2II} & E_{2} & 0 & B_{2II}\sqrt{1-D_{II}^{2}} & 0 & 0 & C_{2II}\sqrt{1-E_{2}^{2}} & 0 \\
B_{3II} & D_{II} & C_{3II} & E_{3} & 0 & B_{3II}\sqrt{1-D_{II}^{2}} & 0 & 0 & 0 & C_{3II}\sqrt{1-E_{1}^{2}} \\
B_{1III} & D_{III} & C_{1III} & E_{1} & 0 & 0 & B_{1}\sqrt{1-D_{III}^{2}} & C_{1III}\sqrt{1-E_{1}^{2}} & 0 & 0 \\
B_{2III} & D_{III} & C_{2III} & E_{2} & 0 & 0 & B_{2}\sqrt{1-D_{III}^{2}} & 0 & C_{3II}\sqrt{1-E_{2}^{2}} & 0 \\
B_{3III} & D_{III} & C_{3III} & E_{3} & 0 & 0 & B_{3}\sqrt{1-D_{III}^{2}} & 0 & 0 & C_{3III}\sqrt{1-E_{1}^{2}}
\end{bmatrix}
$$

How can one estimate the unknowns, B_{jk} , C_{jk} , D_j and E_k , for each method-trait combination? A reasonable procedure appears to be the following.

One may choose to estimate the parameters such that the sum of squares of residuals in the off-diagonal elements of the correlation matrix are minimized. That is, one minimizes Q_1 , where

$$Q_1 = \sum_{q<q'} (A_{qq'} - \hat{\Sigma}_{qq'})^2 \qquad q,q' = 1,2, \ldots ,p \, .$$

A = observed correlation matrix

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda} \, .$$

The function Q_1 is the same type of function minimized by Harman (1966) in exploratory research on optimization procedures in factor analysis.

Or, one may choose to minimize a function analogous to that minimized in maximum likelihood factor analysis. That is, one minimizes Q_2 , where

$$Q_2 = \sum_{q<q'} (A_{qq'} - \hat{\Sigma}_{qq'})^2/\hat{\psi}_{qq} \, \hat{\psi}_{q'q'} \text{ and } \hat{\Sigma} = \hat{\Lambda}\hat{\Lambda} + \hat{\psi} \, .$$

This function is different from the one discussed earlier in the dissertation. The difference occurs in the type of restriction imposed on the factor structure matrix (Λ). That is, the restrictions required here are not the same type as those imposed in the restricted maximum likelihood factor analysis. In addition to requiring that certain elements be zero, one requires that the elements be multiplicative entities. The RMLFA Computer Program is not readily amenable to incorporation of restrictions of this type.

The two functions, Q_1 and Q_2, are of interest for a number of reasons. The function Q_1 weights all residuals equally. Results of minimization of this function may approximate those obtained using the usual likelihood function. The degree of approximation depends on the variability of the specific variances and the size of the residuals. If the specific variances are homogeneous the residuals are small, then the approximation will be a good one. However, Q_1 may be the more appropriate function to minimize in some cases. If some of the model assumptions are not met and the nature of the violation is not clear, then weighting residuals differentially may not be tenable. That is, Q_1 may be used insofar as one is ignorant of a more appropriate function, and Q_2 is inappropriate because of violated assumptions.

An example of such a situation is given by observed variables whose distributions are not normal. Tests of very high or very low difficulty yield skewed distributions. Variables whose distributions are skewed in the same direction are more likely to correlate highly with each other than with variables whose distributions are skewed in the opposite direction. Insofar as differential weighting of residuals attaches more relative importance to high correlations, the results will be misleading.

In order to minimize either of the functions, Q_1 or Q_2, one can use an optimization procedure such as the method of parallel tangents or PARTAN. The PARTAN algorithm is described in Shah, Buehler and Kempthorne (1964) and in Buehler, Shah and Kempthorne (1964). The method allows one to attack the general problem of determining values of the variables X_1, X_2, ... ,X_n, which minimize or maximize the function $Q(X_1,X_2, \ldots ,X_n)$. PARTAN requires that the derivatives with respect to each of the unknowns be computed. A convenient mode of presentation of these derivations is in matrix notation, as described by Dwyer (1967). The writer has computed the derivatives associated with Q_1 and Q_2 in the manner prescribed by Dwyer, for the case of three methods and three traits. This system of equations has been programmed in FORTRAN IV as a subroutine (FCN). The subroutines have been tailored for use with the computer program developed by Doerfler (1964) for implementation of the PARTAN algorithm. The relevant information is presented in the Appendix.

B. Method 2

Bock and Bargmann (1966) present a general method for estimating variance components for the case of the one way (random) classification design. This model is related to the factor analytic model in a manner analogous to that shown in this dissertation. Some differences are apparent. Specifically, Bock and Bargmann base their procedure on the following equations, where the definitions of terms have been provided above.

$$y = \Lambda z + s$$

$$E(yy') = E(\Lambda zz'\Lambda) + E(ss')$$

$$\Sigma = \Lambda\Phi\Lambda' + \psi \quad .$$

The usual assumptions of distributional properties are made. However, some unique restrictions are imposed on the factor structure Λ . One assumes or knows the parameters in Λ . This is, of course, contrary to ordinary circumstances insofar as these parameters must usually be estimated. The elements in Λ are restricted to be zero or one. In addition, for the most restrictive case, the authors constrain ψ to γI , where γ is unknown. If Λ is a lower triangular matrix of 1's, and of full rank, then one can show that the factor analytic model is, in some sense, equivalent to the model

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad \alpha_i \sim NID(0,\sigma^2_{\alpha_i})$$

$$\varepsilon_{ij} \sim NID(0,\sigma^2_\varepsilon)$$

$$i = 1,2, \ldots ,a$$

$$j = 1,2, \ldots ,N \quad .$$

One can, using the Bock and Bargmann procedure, estimate variance components for each effect, i.e. $\hat{\sigma}^2_{\alpha_1}, \hat{\sigma}^2_{\alpha_2}, \ldots, \hat{\sigma}^2_{\alpha_2}$ and the homoscedastic error, $\hat{\sigma}^2_\varepsilon$. The estimates are maximum likelihood estimates.

Bock and Bargmann use the Newton-Raphson method of optimization of the likelihood function which they derive.

The approach is relevant insofar as it provides another method of attack on the problem of relationships between the factor analytic and classical methods of analysis. The multimethod-multitrait data appear to be amenable to analysis in the manner discussed. However, the implementation of the analysis is difficult for a number of reasons. The major deterent is the lack of a convenient computer program for the analysis.

VI. SUGGESTIONS FOR FURTHER RESEARCH

A. Applications In Experimental Psychology

Within experimental psychology, problems are evident with respect to (1) response measures and (2) assessment of relations among response measures. Consider, for example, various indices which psychologists have proposed to indicate "arousal". The relations among skin conductance, skin potential, Critical Fusion Flicker, blood pressure, respiratory rate, GSR, self ratings, experimenter ratings, etc. are not well documented. By well documented is meant that such relations are systematically examined under some standard series of conditions and/or subjects and statistical procedures are used to assess the nature of the responses and the relations among them.

Given that such response relations are not well documented, what can one do about this?

A possible approach to systematic examination of the problem is proposed here: multitrait-multimethod paradigm together with the analytic procedures outlined in this research. Some obvious examples are developed. In the following examples, T_j represents the jth trait, and M_k represents method k .

Example 1   Suppose one wishes to assess the influence of various rater biases in a standard situation, supposed to produce some level of arousal. One can examine the situation defined by the following methods and traits.

T_1 - skin conductance	M_1 - Experimenter 1
T_2 - skin potential	M_2 - Experimenter 2
T_3 - finger pulse volume	M_3 - Experimenter 3
T_4 - respiration rate	
T_5 - temperature	

Using the analytic procedure described, based on the correlations among the responses which result from the paradigm above, one can assess several hypothesized factors. Perhaps the most important is method bias as a function of experimenter. Rosenthal (1963) has done some research relevant to experimenter induced bias, which suggests that investigation of this type may be fruitful. The convergent and discriminant validity of the "trait" above can be assessed in terms of arousal. If, in fact, all these do measure arousal there should be high convergent validity, low discriminant validity.

Example 2 On could use various "datum points" rather than experimenters in the example above. The question of where to start or stop observations, based on a continuous response, seems to be of some importance according to, say, Venables and Martin (1967). From the proposed analysis, one might be better able to determine what sort of criteria should be used in order to obtain consistent results.

Example 3 Suppose one wished to examine some common conceptions of drives, characterized as traits, and alleged different. Several methods of measurement may be appropriate for their assessment, for example, those described below.

T_1 - Anxiety

T_2 - Arousal

T_3 - Emotionality

T_4 - Neuroticism

T_5 - Psychoticism

M_1 - MMPI, MAS, Rorschach, etc.

M_2 - Physiological measures

M_3 - Kinesthetic judgements, memory span, visual field, etc.

The analysis would yield some basis for intelligent evaluation of method bias, of degree to which each of these drives may be different (given a controlled situation), the degree to which the drives are not distinguishable from one another.

Example 4 Levels of a particular drive under consideration may be examined, the responses associated with the levels considered as traits. The methods may be of the form in Examples 3 or a subset of these. This might be particularly revealing in the light of sections of the response continuum, assuming that something like the Yerkes-Dodson Law holds for the complete continuum. That is, one considers only those sections of the continuum which are nearly linear, or can be transformed easily. Method bias is again obtained. In addition, one may be able to examine the extent to which the traits, or levels of stimulation, are different entities. That is, the total response may not be a simple additive function of stimulation, but may be more consistent with Selye's ideas on the production of entirely new responses, Selye (1956).

Example 5 Experimental psychologists have only recently begun to seriously consider the problem of individual differences in responses of humans or animals. The paradigm described may be useful in exploring the problem further. Specifically, one might consider the traits to be different response measures (as in the first example), or levels of drive (as in the fourth example). The methods would be subjects, where correlations are computed over time, levels of the trait, situations which are incremental in some respect and which may be relevant. Results of such an analysis should indicate the severity of the individual differences problem for the specified situation (method bias). Some indicators of alternative measurement techniques, or at least, the best measurement techniques should be evident. By best here is meant, those which reflect individual differences least.

B. Applications In Psychometrics

There appears to be some justification for further exploration of the procedures described in this dissertation. Specifically, one might consider the following.

1. There appear to be rather strong relationships between the type of research presented here to the work done by Bock and Bargmann (1966), by Gollob (1968) and by others cited in the introductory section of this dissertation. Each of the efforts is directed, in part, to estimation of parameters in some classificatory models by using factor analytic techniques. The problem of general design models and their relation to the less restrictive factor analytic model deserves further attention.
2. One convenient method of exploring an array of classificatory design models, in a factor analytic sense, is through Monte Carlo techniques. Consider, for example, a multimethod-multitrait matrix which is ideal in some sense. The ideality is based, say, on the known linearity of the

model, the known tenability of distributive assumptions, and large sample size. One can obtain such ideality through generation of random numbers for a multivariate normal where the variables are correlated. A situation such as this may led itself well to more systematic examination of the relation of the classical design model to a factor analytic model with restrictions. At, least one might learn a bit more about interpretation of results.

3. One of the useful features of the maximum likelihood factor analysis is the possibility of testing hypotheses. The χ^2 test developed, however, is applicable only to the case of a single decision. That is, there is no information for the case of sequential testing, conditional on the various models hypothesized. This, of course, is a pervasive problem in statistics and beyond the competencies of this writer. It seems not unreasonable to aspire toward the development of some procedure to make the analysis more rigorous than it is, in a testing sense.

4. The expansion of one of the first of the alternative procedures discussed above may be rewarding. That is, there is a great deal of flexibility inherent in using an algorithm such as PARTAN for optimization of various functions, conditional on a variety of models. The only real difficulties for implementation are computation of derivatives in a form suitable for computer programming applications. Although the questions of unbiasedness, minimum variance, etc. of estimators are important, there seems to be no good reason why more exploratory research cannot be carried out. One may consider the use of numerical methods for obtaining derivatives and Monte Carlo generation of idealized data.

VII. SUMMARY

The procedure described may be summarized from a functional point of view.

Given a number of allegedly different methods of measuring some (other) number of allegedly different attributes or traits of an experimental unit, one can use the analytic procedure to examine specific aspects of this situation. The procedure is essentially a quantification and expansion of the systematic assessment of multitrait-multimethod matrices as described by Campbell and Fiske (1959).

Specific aspects which can be examined are the following.

1. Given a number of methods of measurement, one can assess the extent to which the observation is influenced by the particular method. That is, the results of the analysis provide a means of judging which methods are biased, insofar as the method contributes undesirable variance to the observations. One can do this for particular traits. That is, given two traits, the method bias may enter more heavily into measurement of one than it enters into the measurement of the other.
2. One can examine the alleged discriminability of traits. Moreover, one can do this regardless of which method of measurement is used. Specifically one can establish the degree to which traits or attributes are related to one another after one has accounted for method biases. In the Campbell-Fiske nomenclature this is known as discriminant validation.
3. The statistical procedure requires that one hypothesize a linear model to account for the data. The inferences described in the two items above are, of course, conditional on this model being true. The procedure

developed has the distinct advantage of allowing one to assess the goodness of fit of the model. This is done through the use of a χ^2 statistic.

4. One can also assess the extent to which individual differences contribute to the observations, independent of the particular method-trait combination used in measurement. This is important insofar as one would like to examine global factors such as general reputation, etc. in assessment of individuals.

5. The analytic procedure, which is used to extract the information given above, is based on relations between the factor analytic model and the randomized blocks design. The technique allows one to assess the extent to which various assumptions made in analysis of variance are met. Specifically, one might wish to examine main effects, interaction and error with respect to homogeneity of variance and correlations between the independent variables.

VIII. LITERATURE CITED

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley and Sons, 1958.

Anderson, T. W. and Rubin, H. Statistical inference in factor analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1956, 111-150.

Anscombe, F. J. Examination of residuals. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1960, 1, 1-36.

Bartlett, M. S. The effect of standardization on an approximation in factor analysis. Biometrika, 1951, 38, 337-344.

Bock, R. D. and Bargmann, R. Analysis of covariance structures. Psychometrika, 1966, 4, 507-534.

Boneau, C. A. The effects of violations of assumptions underlying the t test. Psychological Bulletin, 1960, 57, 49-64.

Borgatta, E. F. Analysis of social interaction: Actual, role-playing and projective. Journal of Abnormal and Social Psychology, 1955, 51, 394-405.

Boruch, R. F. and Wolins, L. Factor analysis with constraints (or analysis of variance with fewer constraints). Paper presented at Statistics Seminar, Iowa State University, Ames, 1968.

Buehler, R. J., Shah, B. V. and Kempthorne, O. Methods of parallel tangents. Chemical Engineering Progress symposium Series, 1964, 60, 1-7.

Campbell, D. T. Direct iterative factoring in matrices with crossed factor-classes. Unpublished Manuscript, Northwestern University, 1965.

Campbell, D. T. and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Campbell, D. T. and O'Connell, E. J. Method factors in multimethod-multitrait matrices: multiplicative rather than additive? Unpublished Manuscript, Northwestern University, 1965.

Cureton, E. E. Unbiased estimation of the standard deviation. American Statistician, 1968, 22, 22.

Daniel, C. Study of a multiple regression. Unpublished paper presented at Applied Statistics Seminar, Iowa State University, Ames, 1968.

Doerfler, Thomas E. Fortran subroutine: PARTAN minimization by the method of parallel tangents. Unpublished Manuscript, Iowa State University, 1964.

Dwyer, P. S. Some applications of matrix derivatives in multivariate analysis. Journal of the American Statistical Association, 1967, 62, 607-625.

Einstein, A. The world as I see it. New York: Covici Friede, 1934.

Fletcher, R. and Powell, M. J. D. A rapidly convergent descent method for minimization. Computer Journal, 1963, 2, 163-168.

Gollob, H. F. A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika, 1968, 33, 73-115.

Gulliksen, H. Theory of mental tests. New York: Wiley and Sons, 1950.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Harman, H. H. and Jones, W. H. Factor analysis by minimizing residuals (MINRES). Psychometrika, 1966, 31, 351-368.

Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart and Winston, Incorporated, 1965.

Joreskog, K. G. RMLFA, a computer program for restricted maximum likelihood factor analysis. Research Bulletin RB-67-21, Educational Testing Service, Princeton, New Jersey, 1967a.

Joreskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1967b, 32, 443-482.

Kaplan, Abraham. The conduct of inquiry. San Francisco: Chandler Publishing Company, 1964.

Kelly, E. L. and Fiske, D. W. The prediction of performance in clinical psychology. Ann Arbor: University of Michigan Press, 1951.

Kempthorne, O. The design and analysis of experiments. New York: John Wiley and Sons, 1952.

Kempthorne, O. Non-linear models. Unpublished paper presented at Applied Statistics Seminar, Iowa State University, Ames, 1967.

Lana, R. E. and Lubin, A. The effect of correlation on the repeated measurements design. Educational and Psychological Measurements, 1963, 4, 729-739.

Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, Section A, 1940, 60, 64-82.

Lawley, D. N. and Maxwell, A. E. Factor analysis as a statistical method. London: Butterworth. 1963.

MacKinney, A. C. The longitudinal study of managerial performance. Paper presented at the Annual Meetins, Midwest Business Administration Association, Minneapolis, April 18, 1968.

Morrison, D. F. Multivariate statistical methods. New York: McGraw Hill, 1967.

Rao, C. R. Estimation and tests of significance in factor analysis. Psychometrika, 1955, 20, 93-111.

Rosenthal, R. On the social psychology of the psychological experiment: the experimenters' hypothesis as unintended determinant of experimental results. American Scientist, 1963, 51, 268-282.

Selye, H. The stress of life. New York: McGraw Hill, 1956.

Shah, B. V., Buehler, R. J. and Kempthorne, O. Some algorithms for minimizing a function of several variables. Journal of Industrial and Applied Mathematics, 1964, 12, 74-92.

Stanley, J. C. Fixed, random and mixed models in the analysis of variance as special cases of a finite model. Psychological Reports, 1956, 2, 369.

Stanley, J. C. Analysis of unreplicated three way classifications, with applications to rater bias and trait independence. Psychometrika, 1961, 26, 205-219.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Tucker, Ledyard R. The extension of factor analysis to three dimensional matrices. In N. Frederiksen and H. Gulliksen (Eds.), Contributions to mathematical psychology. New York: Holt, Rinehart and Winston, 1964, 109-127.

Tucker, Ledyard R. Some mathematical notes on three mode factor analysis. Unpublished Manuscript, University of Illinois, 1965.

Venables, P. H. and Martin, Irene. A manual of psychophysical methods. New York: Wiley and Sons, 1967.

Walker, Helen M. and Lev, J. Statistical inference. New York: Holt, Rinehart and Winston, 1953.

Wolins, L. A procedure for estimating the amount of trait, method and error variance attributable to a measure. Project C-998, Contract 3-20-001, Office of Education, U.S. Department of Health, Education and Welfare: (Donald T. Campbell, Principle Investigator), 1964.

Zyzanski, S. J. Analysis of variance applied to factors which do not have comparable scales. Unpublished M.S. thesis, Iowa State University of Science and Technology, 1962.

IX. ACKNOWLEDGEMENTS

I wish to express my appreciation to all those who have been helpful in the preparation of this dissertation. Professor Leroy Wolins has generously contributed the ideas and skills which made initiation of this research possible. I should like to thank the members of my committee for their interest, attention and, occasionally, forebearance.

APPENDIX A

1. Minimization of $Q_1 = \sum_{j \neq k} (A_{jk} - \hat{\Sigma}_{jk})^2$, $j,k = 1,2,3$.

A = 9 x 9 matrix of observed correlations

$\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}$ = 9 x 9 matrix of estimated correlations

$\hat{\Lambda}$ = 9 x 8 matrix of factor loadings.

For convenience of derivation and computer programming let the Λ matrix be defined as in Table 19. An element in the matrix above is represented as x_i , $i = 1,2, \ldots ,24$, a parameter to be estimated.

Each element in the Λ matrix shown, corresponds to a variable with respect to which, one must take the derivative. There is a one to one correspondence between this matrix and the one described earlier in the section on alternative methods.

In matrix notation,

$$Q_1 = \text{trace } \{[(A-I) - (\Lambda\Lambda'\text{-Diag } \Lambda\Lambda')]'[(A-I) - (\Lambda\Lambda'\text{-Diag } \Lambda\Lambda')]\} .$$

Q_1 is of the form $Q = \text{trace } (Y'AY)$ and, in general,

$$\frac{\partial Q}{\partial y_{jk}} = \text{trace } \frac{\partial}{dy_{jk}} (Y'AY) = \text{trace } (J'AY + Y'AJ)$$

where

$$J = \frac{\partial Y}{\partial y_{jk}}$$ and y_{jk} is a function of the jkth element of an arbitrary matrix Y .

Substituting,

$$j = \frac{\partial Y}{\partial y_{jk}} = \frac{\partial}{\partial X_i}[(A-I) - (\Lambda\Lambda'\text{-Diag } \Lambda\Lambda')] = (0-0) - \frac{\partial}{\partial X_i} (\Lambda\Lambda'\text{-Diag } \Lambda\Lambda') \quad .$$

Table 19. Λ matrix

$$\Lambda = \begin{bmatrix}
x_1x_{10} & x_{13}x_{22} & x_1\sqrt{1-x_{10}^2} & 0 & 0 & x_{13}\sqrt{1-x_{32}^2} & 0 & 0 \\
x_2x_{10} & x_{14}x_{23} & x_2\sqrt{1-x_{10}^2} & 0 & 0 & 0 & x_{14}\sqrt{1-x_{23}^2} & 0 \\
x_3x_{10} & x_{15}x_{24} & x_3\sqrt{1-x_{10}^2} & 0 & 0 & 0 & 0 & x_{15}\sqrt{1-x_{24}^2} \\
x_4x_{11} & x_{16}x_{22} & 0 & x_4\sqrt{1-x_{11}^2} & 0 & x_{16}\sqrt{1-x_{22}^2} & 0 & 0 \\
x_5x_{11} & x_{17}x_{23} & 0 & x_5\sqrt{1-x_{11}^2} & 0 & 0 & x_{17}\sqrt{1-x_{23}^2} & 0 \\
x_6x_{11} & x_{18}x_{24} & 0 & x_6\sqrt{1-x_{11}^2} & 0 & 0 & 0 & x_{18}\sqrt{1-x_{24}^2} \\
x_7x_{12} & x_{19}x_{22} & 0 & 0 & x_7\sqrt{1-x_{12}^2} & x_{19}\sqrt{1-x_{22}^2} & 0 & 0 \\
x_8x_{12} & x_{20}x_{23} & 0 & 0 & x_8\sqrt{1-x_{12}^2} & 0 & x_{20}\sqrt{1-x_{23}^2} & 0 \\
x_9x_{12} & x_{21}x_{24} & 0 & 0 & x_9\sqrt{1-x_{12}^2} & 0 & 0 & x_{21}\sqrt{1-x_{24}^2}
\end{bmatrix}$$

Now,

$\frac{\partial}{\partial X_i} \Lambda\Lambda'$ is of the form $\frac{\partial Q}{\partial y_{jk}} = \frac{\partial}{\partial y_{jk}} YAY' = J'AY' + YAJ'$,

$\frac{\partial \Lambda\Lambda'}{\partial X_i} = J_\Lambda \Lambda' + \Lambda J_\Lambda'$ and $J_\Lambda = \frac{\partial \Lambda}{\partial X_i}$,

$$\frac{\partial}{\partial X_i} \text{Diag } \Lambda\Lambda' = \text{Diag } \frac{\partial}{\partial X_i} \Lambda\Lambda' = \text{Diag } (J_\Lambda \Lambda' + \Lambda J_\Lambda') = D \ .$$

Substituting,

$$j = \frac{\partial}{\partial X_i} [(A-I) - (\Lambda\Lambda' - \text{Diag } \Lambda\Lambda')]$$

$$= [(0-0) - [(J_\Lambda \Lambda' + \Lambda J_\Lambda') - D] \ .$$

Also, one can prove that $\frac{\partial Y}{\partial y_{jk}} = [\frac{\partial Y'}{\partial y_{jk}}]'$.

Using these equalities in the original equation for Q_1 we obtain the following general expression for the derivative with respect to any unknown.

$$\frac{\partial Q_1}{\partial X_i} = (-1) \text{ trace } \{[(J_\Lambda \Lambda' + \Lambda J_\Lambda') - D]'[(A-I) - (\Lambda\Lambda' - \text{Diag } \Lambda\Lambda')]$$
$$+ [(A-I) - (\Lambda\Lambda' - \text{Diag } \Lambda\Lambda')]'[(J_\Lambda \Lambda' + \Lambda J_\Lambda') - D]\} \ .$$

2. Minimization of $Q_2 = \sum_{j,k}(A_{jk} - \hat{\psi}_{jk} - \hat{\Sigma}_{jk})^2/\psi_{jj}\psi_{kk}$.

A = p x p matrix of observed correlations

$\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda} + \hat{\psi}$ = p x p matrix of estimated correlations

$\hat{\Lambda}$ = p x m matrix of factor loadings

$\hat{\psi}$ = p x p matrix of specific variances (diagonal).

For convenience of derivation and computer programming let the variables in Λ be represented as in the case of Q_1 above. That is, the parameters in Λ which must be estimated are symbolized by x_i , $i = 1,2, \ldots ,24$. In addition, one must estimate the elements in ψ . Let these be symbolized by x_i , $i = 25,26, \ldots ,33$.

In matrix notation,

$$Q_2 = \text{trace } [(\psi^{-1})(A-\psi-\Lambda\Lambda')'(\psi^{-1})(\Lambda-\psi-\Lambda\Lambda')] \quad .$$

Q_2 is of the form trace (QZ) where Q and Z are different matrix functions of, say, y_{jk} . In general

$$\frac{\partial QZ}{\partial y_{jk}} = QJ_Z + J_Q Z \quad .$$

Substituting, one obtains the following.

$$\frac{\partial}{\partial X_i} [(\psi^{-1})(A-\psi-\Lambda\Lambda')'][(\psi^{-1})(A-\psi-\Lambda\Lambda')]$$

$$= [(\psi^{-1})(A-\psi-\Lambda\Lambda')'][\frac{\partial}{\partial X_i} (\psi^{-1})(A-\psi-\Lambda\Lambda')]$$

$$+ [\frac{\partial}{\partial \psi_i} [(\psi^{-1})(A-\psi-\Lambda\Lambda')']][(\psi^{-1})(A-\psi-\Lambda\Lambda')] \quad .$$

Taking each of the terms separately one obtains the following.

$$\frac{\partial}{\partial X_i} [(\psi^{-1})(A-\psi-\Lambda\Lambda')] = (\psi^{-1})[\frac{\partial}{\partial X_i} (A-\psi-\Lambda\Lambda')]$$

$$+[\frac{\partial}{\partial X_i} \psi^{-1}](A-\psi-\Lambda\Lambda') \quad .$$

$$\frac{\partial}{X_i} [(\psi^{-1})(A-\psi-\Lambda\Lambda')'] = (\psi^{-1})[\frac{\partial}{\partial X_i}(A-\psi-\Lambda\Lambda')']$$

$$= [\frac{\partial}{\partial X_i} (\psi^{-1})](A-\psi-\Lambda\Lambda')' \quad .$$

If $i = 1,2, \ldots ,24$, then

$$\frac{\partial}{\partial X_i} (A-\psi-\Lambda\Lambda') = \frac{\partial\Lambda\Lambda'}{\partial X_i} = J_\Lambda\Lambda' + \Lambda J_\Lambda' \text{ , } J_\Lambda = \frac{\partial}{\partial X_i} \Lambda \text{ .}$$

If $i = 2,5,26, \ldots ,33$, then

$$\frac{\partial}{\partial X_i} (A-\psi-\Lambda\Lambda') = 0 - \frac{\partial\psi}{\partial X_i} - 0 = K_\psi \text{ .}$$

$\{K_\psi\}_{jk} = \delta_{jk}$ where δ_{jk} is the kronecker delta.

$$\frac{\partial}{\partial X_i} (\psi^{-1}) = K_{\psi^{-1}} \text{ .}$$

The derivatives with respect to x_i , $i = 1,2, \ldots ,24$ are represented by the expression

$$\frac{\partial Q_2}{\partial X_i} = (-1) \text{ trace } \{[(\psi^{-1})(A-\psi-\Lambda\Lambda')'][(\psi^{-1})(J_\Lambda\Lambda' + \Lambda J_\Lambda')]$$
$$+ [(\psi^{-1})(J_\Lambda\Lambda' + \Lambda J_\Lambda')][(\psi^{-1})(A-\psi-\Lambda\Lambda')]\} \text{ .}$$

The derivatives with respect to x_i, $i = 25, 26, \ldots ,33$ are represented by the expression

$$\frac{\partial Q_2}{\partial X_i} = \text{trace } [(\psi^{-1})(A-\psi-\Lambda\Lambda')'][(\psi^{-1})(-K_{\psi'}) + (K_{\psi^{-1}})(A-\psi-\Lambda\Lambda')]$$
$$+ [(\psi^{-1})(-K_\psi) + (K_{\psi-1})(A-\psi-\Lambda\Lambda')'][(\psi^{-1})(A-\psi-\Lambda\Lambda')]\} \text{ .}$$

These representations are convenient for computer programming.

APPENDIX B

The two simple computer programs presented are written in FORTRAN IV for use on the IBM 360-65. The programs implement the analytic procedure derived in Appendix A, and contain comment cards to simplify interpretation.

The first program given is a subroutine called FCN, tailored for use with the optimization program devised by Doerfler (1964). The program computes the values of first derivatives and values of the function Q_1, described earlier, at the various points examined for an optimum.

The second program presented is a test program in conjunction with subroutine FCN devised for the case of optimization of the weighted residuals function, Q_2. The test program is used as a check on the computations of the subroutine. The test program and test cards in the subroutine should be removed prior to use of the subroutine FCN with PARTAN.

```
      SUBROUTINE FCN(I,Y,X,G)
C
C
C     UNWEIGHTED RESIDUALS CASE
C
C     SUBROUTINE COMPUTES THE VALUE OF FUNCTION Y, VALUE OF
C     DERIVATIVE G, AT THE POINT X
C
C
      DIMENSION X(50),G(50),FG(9,8),PON(9,9),PTN(9,9),
     1RH(9,9),F(9,8),PTH(9,9),PF(9,9),PTW(9,9),SUM(9,9)
      DOUBLE PRECISION X,G,FG,PON,PTN,SUM,  RH,F,PTH,PF,PTW
      DOUBLE PRECISION R(20,20)
      COMMON ML,NCAL
      COMMON R
C
C
C
C
C
C
C     VALUE OF F MATRIX
C
C
      DO 1 II=1,9
      DO 1 J=1,8
    1 F(II,J)=0.
      K1=1
      K2=3
      K3=3
C
      DO  5  J=10,12
      DO 10 II=K1,K2
      IF(II-10) 4,999,999
  999 STOP 9998
    4 CONTINUE
      F(II,1)=(X(II))*(X(J))
      F(II,K3)=(X(II))*DSQRT(1.0-(X(J))*(X(J)))
   10 CONTINUE
      K3=K3+1
      K2=K2+3
      K1=K1+3
    5 CONTINUE
C
      K2=13
      K3=19
      K4=6
      K5=22
      DO 9 J=1,3
      DO 11 II=K2,K3,3
      IF(K3-22) 13,12,13
```

```
   13 CONTINUE
      K1=II-12
      F(K1,2)= (X(II))*(X(K5))
      F(K1,K4)=(X(II))*DSQRT(1.0-(X(K5))*(X(K5)))
   11 CONTINUE
      K5=K5+1
      K4=K4+1
      K3=K3+1
      K2=K2+1
    9 CONTINUE
   12 CONTINUE
      DO 6 II=1,9
    6 R(II,II)=0.
      WRITE(3,21)
   21 FORMAT(@0 INPUT F MATRIX @)
      DO 22  II=I,9
   22 WRITE(3,23) (F(II,J),J=1,8)
   23 FORMAT(1H0,10F10.4)
C
C
      WRITE(3,24)
   24 FORMAT(@0  INPUT R MATRIX@)
      DO 25 II=1,9
   25 WRITE(3,23) (R(II,J),J=1,9)
C
C     DERIVATIVES WITH RESPECT TO X(1),X(2),...X(9).
C
  100 CONTINUE
      DO 36 L=1,I
      DO 101 J=1,9
      DO 101 K=1,8
  101 FG(J,K)=0.0D0
      IF(L-3)   201,201,202
  201 FG(L,1)=X(10)
      FG(L,3)=DSQRT(1.-(X(10))*(X(10)))
      GO TO 300
  202 IF(L-6)   203,203,204
  203 FG(L,1)=X(11)
      FG(L,3)=DSQRT(1.-(X(11))*(X(11)))
      GO TO 300
  204 IF(L-9)   205,205,206
  205 FG(L,1)=X(12)
      FG(L,3)=DSQRT(1.-(X(12))*(X(12)))
      GO TO 300
C
C     GO TO 300 IS COMPLETION OF DERIVATIVE...G(I).
C
C     DERIVATIVE WITH RESPECT TO X(10),X(11),X(12).
C
  206 IF(L-12)   207,207,208
  207 K2=(L-9)*3
```

```
      K1=K2-2
      DO 210 K=K1,K2
      M9=L-7
      FG(K,1) = X(K)
  210 FG(K,M9)=(-X(K))*(X(L))/(DSQRT(1.-(X(L))*(X(L)) ))
      GO TO 300
C
C     DERIVATIVE WITH RESPECT TO X(13),X(14),...X(21).
C
  208 CONTINUE
      DO 2209  K=13,19,3
      IF(L-K)  2209,211,2209
 2209 CONTINUE
      GO TO 212
  211 K1=L-12
      K2=6
      K3=22
      GO TO 2210
  212 DO 213 K=14,20,3
      IF(L-K)  213,214,213
  213 CONTINUE
      GO TO 215
  214 K1=L-12
      K2=7
      K3=23
      GO TO 2210
  215 DO 216  K=15,21,3
      IF(L-K) 216,217,216
  216 CONTINUE
      GO TO 218
  217 K1=L-12
      K2=8
      K3=24
 2210 CONTINUE
      FG(K1,2)=X(K3)
      FG(K1,K2)=DSQRT(1.0-(X(K3))*(X(K3)))
      GO TO 300
C
C     DERIVATIVE WITH RESPECT TO X(22),X(23),X(24).
C
  218 CONTINUE
      IF(L-22)  219,220,221
  220 K1=13
      K2=19
      J1=6
      GO TO 229
  219 STOP 9999
  221 IF(L-23) 219,223,224
  223 K1=14
      K2=20
      J1=7
```

```
      GO TO 229
  224 IF(L-24)  219,225,219
  225 K1=15
      K2=21
      J1=8
      GO TO 229
  229 DO 230 K=K1,K2,3
      KK=K-12
      FG(KK,2)=X(K)
  230 FG(KK,J1)=(-X(K))*(X(L))/(DSQRT(1.0-(X(L))*(X(L))))
      GO TO 300
C
C     COMPLETION OF DERIVATIVE
C
  300 CONTINUE
      DO 301 II=1,9
      DO 301  J=1,9
      PON(II,J)=0.0D0
      PTW(II,J)=0.0D0
      DO 301  K=1,8
      PON(II,J)=PON(II,J)+(FG(II,K))*(F(J,K))
  301 PTW(II,J)=PTW(II,J)+(F(II,K))*(FG(J,K))
C
C     ABOVE ARE FGF@ AND FFG@ RESPECTIVELY
C
      DO 303  II=1,9
      DO 303   J=1,9
  303 SUM(II,J)=PON(II,J)+PTW(II,J)
      DO 304  J=1,9
  304 SUM(J,J)=0.0D0
      DO 305  J=1,9
  305 R(J,J)=0.0D0
      DO 306 II=1,9
      DO 306  J=1,9
      RH(II,J)=0.0D0
      DO 306  K=1,8
  306 RH(II,J)=RH(II,J)+(F(II,K))*(F(J,K))
      DO 307  J=1,9
  307 RH(J,J)=RH(J,J)-RH(J,J)
      DO 308  II=1,9
      DO 308   J=1,9
      PTH(II,J)=0.0D0
      DO 308   K=1,9
  308 PTH(II,J)=PTH(II,J)+ ( SUM(K,II))*( R(K,J)-RH(K,J))
      TRC=0.0D0
      DO 309  J=1,9
  309 TRC=TRC+PTH(J,J)
      TRC=-(2.0)*(TRC)
      G(L)=TRC
      WRITE(3,34)
   34 FORMAT(@0 VALUE OF DERIVATIVE @)
```

```
      WRITE(3,23) G(L)
   36 CONTINUE
C
C     VALUE OF THE FUNCTION Y
C
C
      DO 310 II=1,9
      DO 310  J=1,9
      PF(II,J)=0.0D0
      DO 310  K=1,9
  310 PF(II,J)=PF(II,J)+ (  R(K,II)-RH(K,II))*(R(K,J)-RH(K,J))
C
      Y=0.0D0
      DO 311 J=1,9
  311 Y=Y + PF(J,J)
      WRITE(3,35)
   35 FORMAT(@0 VALUE OF THE FUNCTION @)
      WRITE(3,23) Y
C
      RETURN
      END
```

T.

```
C
C     WEIGHTED RESIDUALS CASE WITH TESING PROGRAM
C
C
C     INPUT FOR TESTING
      DIMENSION X(50),G(50)
      DIMENSION R(9,9)
      DOUBLE PRECISION X,G
        DOUBLE PRECISION R
      COMMON R
      DO 13  I=1,33
      CAT = I
   13 X(I)= CAT/34
      WRITE(3,1)
    1 FORMAT(@0 TEST-X(1), X(2),...X(24) @)
      WRITE(3,2) (X(I) ,I=1,33)
      WRITE(2,777) (X(I),I=1,24)
  777 FORMAT(4D20.3)
  778 FORMAT(9F8.4)
       DO 34 I=1,33
   34 WRITE(2,778) (R(I,J),J=1,33)
    2 FORMAT(1H0,23F5.2)
      I=33
      CALL FCN(I,Y,X,G)
      WRITE(3,33) ((R(I,J),J=1,9),I=1,9)
   33 FORMAT(1H0,9F10.4)
      STOP
      END
```

```
      SUBROUTINE FCN(I,Y,X,G)
C
C     TESTING - C ADDED TO CARDS 14A,15A MUST BE REMOVED
C     BEFORE ACTUAL OPTIMIZATION OF FUNCTION IS IMPLEMENTED
C
C     ALSO COMPUTATION OF R MUST BE DELETED BEFORE ACTUAL
C     OPTIMIZATION
C
C     SUBROUTINE COMPUTES THE VALUE OF FUNCTION Y, VALUE OF
C     DERIVATIVE G, AT POINT X
C
C
      DIMENSION X(50),G(50),FG(9,8),PON(9,9),PTN(9,9),
     1RH(9,9),F(9,8),PTH(9,9),PF(9,9),PTW(9,9),SUM(9,9)
      DIMENSION W(9,9),WI(9,9),DEL(9,9),WK(9,9),WIK(9,9)
      DIMENSION AA(9,9),AB(9,9),AC(9,9),AD(9,9),AE(9,9)
      DIMENSION AH(9,9),AF(9,9),AG(9,9)
      DOUBLE PRECISION R(20,20)
      DOUBLE PRECISION AA,AB,AC,AD,AE,AF,AG,AH
      DOUBLE PRECISION W,WI,DEL,WK,WIK
      DOUBLE PRECISION X,G,FG,PON,PTN,SUM,  RH,F,PTH,PF,PTW
      COMMON R
C
C
C
C
C     DO 3  II=1,9
C   3 READ(1,2) (R(II,J),J=1,9)
C
C   2 FORMAT(9F8.4)
C     VALUE OF F MATRIX
C
C
      DO 1 II=1,9
      DO 1 J=1,8
    1 F(II,J)=0.
      DO 70 II=1,9
      DO 70  J=1,9
      WK(II,J) = 0.0
      WIK(II,J) = 0.0
      W(II,J)=0.0
      WI(II,J)=0.0
   70 DEL(II,J)=0.0
      DO 71  II=1,9
      K=II+24
      W(II,II)=X(K)
   71 WI(II,II)=1/(X(K))
      K1=1
      K2=3
      K3=3
C
```

```
      DO  5  J=10,12
      DO 10 II=K1,K2
      IF(II-10) 4,5,4
    4 CONTINUE
      F(II,1)=(X(II))*(X(J))
      F(II,K3)=(X(II))*DSQRT(1.0-(X(J))*(X(J)))
   10 CONTINUE
      K3=K3+1
      K2=K2+3
      K1=K1+3
    5 CONTINUE
C
      K2=13
      K3=19
      K4=6
      K5=22
      DO 9 J=1,3
      DO 11 II=K2,K3,3
      IF(K3-22) 13,12,13
   13 CONTINUE
      K1=II-12
      F(K1,2)= (X(II))*(X(K5))
      F(K1,K4)=(X(II))*DSQRT(1.0-(X(K5))*(X(K5)))
   11 CONTINUE
      K5=K5+1
      K4=K4+1
      K3=K3+1
      K2=K2+1
    9 CONTINUE
   12 CONTINUE
C
C
C     TEST ADD DOWN
      DO 3001 II=1,9
      DO 3001 J=1,9
      R(II,J)=0
      DO 3001 K=1,8
 3001 R(II,J) = R(II,J) + (F(II,K))*(F(J,K))
      DO 6 II=1,9
    6 R(II,II)=0.
      WRITE(3,21)
   21 FORMAT(@0 INPUT F MATRIX @)
      DO 22  II=1,9
   22 WRITE(3,23) (F(II,J),J=1,8)
   23 FORMAT(1H0,10F10.4)
      WRITE(3,24)
   24 FORMAT(@0  INPUT R MATRIX@)
      DO 25 II=1,9
   25 WRITE(3,23) (R(II,J),J=1,9)
C
C     TEST ADD UP
```

```
C      DERIVATIVES WITH RESPECT TO X(1),X(2),...X(9).
  100 CONTINUE
      DO 36 L=1,I
      DO 101 J=1,9
      DO 101 K=1,8
  101 FG(J,K)=0.0D0
      IF(L-3)    201,201,202
  201 FG(L,1)=X(10)
      FG(L,3)=DSQRT(1.-(X(10))*(X(10)))
      GO TO 300
  202 IF(L-6)    203,203,204
  203 FG(L,1)=X(11)
      FG(L,3)=DSQRT(1.-(X(11))*(X(11)))
      GO TO 300
  204 IF(L-9)    205,205,206
  205 FG(L,1)=X(12)
      FG(L,3)=DSQRT(1.-(X(12))*(X(12)))
      GO TO 300
C
C      GO TO 300 IS COMPLETION OF DERIVATIVE...G(I).
C
C      DERIVATIVE WITH RESPECT TO X(10),X(11),X(12).
  206 IF(L-12)    207,207,208
  207 K2=(L-9)*3
      K1=K2-2
      DO 210 K=K1,K2
      ML=L-7
      FG(K,1) = X(K)
  210 FG(K,ML)=(-X(K))*(X(L))/(DSQRT(1.-(X(L))**2.0))
      GO TO 300
C
C      DERIVATIVE WITH RESPECT TO X(13),X(14),...X(21).
C
  208 CONTINUE
      DO 2209  K=13,19,3
      IF(L-K)  2209,211,2209
 2209 CONTINUE
      GO TO 212
  211 K1=L-12
      K2=6
      K3=22
      GO TO 2210
  212 DO 213 K=14,20,3
      IF(L-K)  213,214,213
  213 CONTINUE
      GO TO 215
  214 K1=L-12
      K2=7
      K3=23
      GO TO 2210
  215 DO 216  K=15,21,3
```

```
      IF(L-K) 216,217,216
  216 CONTINUE
      GO TO 218
  217 K1=L-12
      K2=8
      K3=24
 2210 CONTINUE
      FG(K1,2)=X(K3)
      FG(K1,K2)=DSQRT(1.0-(X(K3))*(X(K3)))
      GO TO 300
C
C     DERIVATIVE WITH RESPECT TO X(22),X(23),X(24).
C
  218 CONTINUE
      IF(L-22)  219,220,221
  220 K1=13
      K2=19
      J1=6
      GO TO 229
  219 STOP 9999
  221 IF(L-23) 219,223,224
  223 K1=14
      K2=20
      J1=7
      GO TO 229
  224 IF(L-24) 219,225,72
  225 K1=15
      K2=21
      J1=8
      GO TO 229
  229 DO 230 K=K1,K2,3
      KK=K-12
      FG(KK,2)=X(K)
  230 FG(KK,J1)=(-X(K))*(X(L))/(DSQRT(1.0-(X(L))*(X(L))))
      GO TO 300
C     DERIVATIVE WITH RESPECT TO X(25)...X(33).
   72 CONTINUE
      WK(L,L)=1.0
      KA=L+24
      WIK(L,L)=1.0/(X(KA))*(X(KA))
      GO TO 300
C
C     COMPLETION OF DERIVATIVE
C     DERIVATIVE OF FUNCTION DESCRIPTION  IS BE FOUND IN THESIS
C
  300 CONTINUE
      DO 301 II=1,9
      DO 301  J=1,9
      PON(II,J)=0.0D0
      PTW(II,J)=0.0D0
      DO 301  K=1,8
```

```
      PON(II,J)=PON(II,J)+(FG(II,K))*(F(J,K))
  301 PTW(II,J)=PTW(II,J)+(F(II,K))*(FG(J,K))
      WRITE(3,26)
   26 FORMAT(@0 PON MATRIX@)
      DO 27  II=1,9
   27 WRITE(3,23) (PON(II,J),J=1,9)
      WRITE(3,28)
   28 FORMAT(@0 PTW MATRIX@)
      DO 29 II=1,9
   29 WRITE(3,23) (PTW(II,J),J=1,9)
      WRITE(3,30)
   30 FORMAT(@0  DERIVATIVE @)
      DO 31  II=1,9
   31 WRITE(3,23) (FG(II,J),J=1,8)
C
C     ABOVE ARE FGF@ AND FFG@ RESPECTIVELY
C
      DO 303  II=1,9
      DO 303   J=1,9
  303 SUM(II,J)=PON(II,J)+PTW(II,J)
C     DO 305  J=1,9
C 305 R(J,J)=0.0D0
      DO 306 II=1,9
      DO 306  J=1,9
      RH(II,J)=0.0D0
      DO 306  K=1,8
  306 RH(II,J)=RH(II,J)+(F(II,K))*(F(J,K))
      IF (L-24) 85,85,86
   85 CONTINUE
      DO 87 II=1,9
      DO 87 J=1,9
   87 DEL(II,J) = R(II,J) - W(II,J) - RH(II,J)
      DO 88 II=1,9
      DO 88 J=1,9
      AA(II,J) = 0.
      AD(II,J)=0.
      AB(II,J) = 0.
      AE(II,J) = 0.
      DO 88 K=1,9
      AA(II,J) = AA(II,J) +(WI(II,K))*(DEL(J,K))
      AB(II,J)=AB(II,J) + (WI(II,K))*(DEL(K,J))
      AD(II,J)=AD(II,J) + (WI(II,K))*(SUM(K,J))
      AE(II,J) = AE(II,J) + (WI(II,K)) *(SUM(J,K))
   88 CONTINUE
      DO 89 II=1,9
      DO 89 J=1,9
      AF(II,J) = 0.
      AG(II,J) = 0.
      DO 89 K=1,9
      AF(II,J) = AF(II,J) + (AA(II,K))*(AD(K,J))
   89 AG(II,J) = AG(II,J) + (AE(II,K))*(AB(K,J))
```

```
      DO 90 J=1,9
   90 PTH(J,J) = AF(J,J) + AG(J,J)
      GO TO 91
C
C     WI*(R-W-FF@)@,WI*(R-W-FF@),WI*WK,WIK*(R-W-FF@),WIK*
C    1(R-W-FF@)@
C
   86 CONTINUE
      DO 74 II=1,9
      DO 74  J=1,9
   74 DEL(II,J)= R(II,J)-W(II,J)-RH(II,J)
      DO 75  II=1,9
      DO 75   J=1,9
      AA(II,J)=0.0
      AB(II,J)=0.0
      AC(II,J)=0.0
      AD(II,J)=0.0
      AE(II,J)=0.0
      DO 75  K=1,9
      AA(II,J)=AA(II,J) +(WI(II,K))*(DEL( J,K))
      AB(II,J)=AB(II,J) +(WI(II,K))*(DEL(K,J))
      AC(II,J)=AC(II,J) +(WI(II,K))*( WK(K,J))
      AD(II,J)=AD(II,J) +(WIK(II,K))*(DEL(K,J))
      AE(II,J)=AE(II,J) +(WIK(II,K))*(DEL(J,K))
   75 CONTINUE
C     COMBINE PARENTHETICAL ELEMENTS WI*WK+WIK*(R-W-FF@) AND
C     WI*WK + WIK*(R-W-FF@)@
      DO 76 II=1,9
      DO 76  J=1,9
      AD(II,J) = AD(II,J) - AC(II,J)
      AE(II,J) = AE(II,J) - AC(II,J)
   76 CONTINUE
      DO 77 II=1,9
      DO 77 J=1,9
      AF(II,J)=0.0
      AG(II,J)=0.0
      DO 77 K=1,9
      AF(II,J)=AF(II,J) +(AA(II,K))*(AD( K,J))
      AG(II,J)=AG(II,J) +(AE(II,K))*(AB( K,J))
   77 CONTINUE
      DO 78  II=1,9
      DO 78  J =1,9
   78 AH(II,J)=AF(II,J) + AG(II,J)
      DO 307  J=1,9
  307 RH(J,J)=RH(J,J)-RH(J,J)
      WRITE(3,32)
   32 FORMAT(@0  RH MATRIX@)
      DO  33 II=1,9
   33 WRITE(3,23) (RH(II,J),J=1,9)
   91 CONTINUE
      IF(L-24) 80,80,81
```

```
   80 CONTINUE
      TRC=0.0D0
      DO 309  J=1,9
  309 TRC=TRC+PTH(J,J)
      GO TO 82
   81 CONTINUE
C
C     DERIVATIVE VALUE WEIGHTED
C
      TRC=0.0D0
      DO 79 J=1,9
   79 TRC = TRC + AH(J,J)
C
C
C     PRODUCT OF WI*DEL@*WI*DEL= Y = AA*AB
C     VALUE OF THE FUNCTION
C
C
C
C
   82 CONTINUE
      DO 83 II=1,9
      DO 83 J=1,9
      PF(II,J) = 0.0
      DO 83 K=1,9
   83 PF(II,J) = PF(II,J) + (AA(II,K))*(AB(K,J))
C
C     VALUE OF THE FUNCTION Y
C
      G(L)=TRC
      Y=0.0D0
      DO 311 J=1,9
  311 Y=Y + PF(J,J)
      WRITE(3,35)
   35 FORMAT(@0 VALUE OF THE FUNCTION @)
      WRITE(3,23) Y
      WRITE(3,34)
   34 FORMAT(@0 VALUE OF DERIVATIVE @)
      WRITE(3,23) G(L)
   36 CONTINUE
      RETURN
      END
```